



UMA INTRODUÇÃO ÀS FILAS DE ESPERA

Cláudia Rossana Velosa Pereira

Mestrado em Matemática
Julho de 2009



UMA INTRODUÇÃO ÀS FILAS DE ESPERA

Cláudia Rossana Velosa Pereira

Tese submetida para obtenção do grau de Mestre em Matemática

Sob orientação de:
Sandra Mendonça, Professora Doutora do Departamento de Matemática e Engenharias da
Universidade da Madeira.

Universidade da Madeira
Julho de 2009

Aos meus pais.

Agradecimentos

Agradeço à minha orientadora Prof^a Doutora Sandra Maria Freitas Mendonça pois esteve sempre disposta a ajudar-me e orientar-me da melhor maneira.

Claro que também tenho a agradecer aos meus colegas e professores, pela sua amizade e apoio. Não posso nomear nomes, pois acabaria por esquecer alguém injustamente. Foi com eles com quem passei o maior tempo da minha vida académica e sem eles tudo iria ser muito mais difícil.

Esta tese é dedicada aos meus pais, pois foi com o seu apoio e dedicação que eu cheguei aqui. Sem eles eu nunca teria conseguido.

Agradeço a todas as pessoas que me apoiaram, que me ajudaram de alguma forma e que sempre acreditaram em mim. Em especial ao meu irmão, Filipe, que me ajudou sempre que precisei. Por último mas muito importante, quero agradecer ao meu namorado, Paulo, pela ajuda que me deu, apoiando-me sempre com o seu amor e dedicação.

A todos, muito obrigado por existirem.

Resumo

Esta tese trata de sistemas de filas de espera estudando o seu comportamento ao longo do tempo e quando se encontram em estado de equilíbrio.

A tese é constituída por três grandes capítulos. Em primeiro lugar são apresentados alguns conceitos básicos da probabilidade, da estatística e de processos de estocásticos. São também descritas as condições e características necessárias para formar um sistema de filas de espera.

Em seguida desenvolvemos vários tipos de sistemas de filas de espera markovianos, estudando várias características de cada modelo, entre elas o número esperado de clientes no sistema e na fila, o tempo esperado que um cliente aguarda no sistema e na fila, após o sistema estar em equilíbrio. Apresentamos também alguns gráficos e comparações.

Por fim, fazemos uma abordagem a alguns sistemas de filas de espera não markovianos, com um estudo menos aprofundado, mas sempre tentando determinar as características que foram determinadas nos modelos markovianos.

Palavras-chave

Sistema de filas de espera, chegadas, fila de espera, serviço, número esperado de clientes, tempo esperado, estado de equilíbrio, fórmulas de Little, processo de nascimento e morte, modelos markovianos, modelos não-markovianos.

Abstract

This thesis talks about systems of waiting queues by studying their behavior along the time and when they are in a state of equilibrium.

The thesis consists of three main chapters. First are presented some basic concepts of probability, statistical and stochastic processes. It also describes the conditions and characteristics necessary to form a system of waiting queues.

Then we develop various types of systems of Markovian waiting queues, studying various characteristics of each model, including the expected number of customers in the queue system and the expected time a customer waits in the queue, after the system is in equilibrium. We also present some graphics and comparisons.

Finally, we present a general approach to some systems of non-Markovian waiting queues, with a less detailed study, but always trying to determine the characteristics that were determined in Markovian models.

Keywords

System of waiting queues, arrivals, waiting queues, service, expected number of customers, expected time, state of equilibrium, Little's formulas, birth and death processes, Markovian models, non-Markovian models.

Conteúdo

1	Introdução	1
2	Conceitos fundamentais	4
2.1	Sigma-álgebra e função mensurável	4
2.2	Independência de acontecimentos	5
2.3	Variável Aleatória	6
2.4	Processos estocásticos	11
2.4.1	Processo de vida e morte	12
3	Teoria das filas de espera	16
3.1	Elementos constituintes de um sistema	16
3.1.1	Chegadas, Fonte ou População	16
3.1.2	Fila de espera	17
3.1.3	Serviço	17
3.1.4	Capacidade do sistema	18
3.1.5	Disciplina de atendimento	18
3.2	Tipos de sistemas de filas de espera	19
3.2.1	Variantes das filas de espera	19
3.3	Características das filas de espera	20
3.3.1	Estado estacionário ou de equilíbrio	21
3.3.2	Algumas notas	21
4	Modelos $M/M/a/b/c$	23
4.1	Modelo $M/M/1$	23
4.2	Modelo $M/M/S$	32
4.3	Modelo $M/M/\infty$	41
4.4	Modelo $M/M/1/K$	44
4.5	Modelo $M/M/S/K$	53
4.5.1	Modelo $M/M/S/S$	62
4.6	Modelo $M/M/1/\infty/H$	66
4.7	Modelo $M/M/S/\infty/H$	72
5	Outros modelos	80
5.1	Modelo $M/E_r/1$	80
5.2	Modelo $M/G/1$	85
5.3	Modelo $M/G/\infty$	93
5.4	Modelo $M/G/S/S$	94
5.5	Modelo $G/M/1$	95
5.6	Modelo $G/G/1$	99
5.7	Modelo $G/G/S$	103
6	Conclusão	106
	Referências	107
	Índice	109

1 Introdução

O estudo das filas de espera surge no trabalho do matemático dinamarquês Agner Krarup Erlang (Lonborg, 1 de Janeiro de 1878 - Lonborg, 3 de Fevereiro de 1929) que, no início do século XX foi a primeira pessoa a estudar o problema das redes de telefones.

Segundo Brockmeyer *et al.* [9] A. K. Erlang começou por estudar a troca de ligações de uma pequena vila e criou uma fórmula, agora conhecida como a fórmula de Erlang, para calcular a fracção de ligações para fora da vila que tinham de esperar porque todas as linhas estavam ocupadas.

Embora o modelo de Erlang seja simples, o seu modelo ainda serve de base à matemática que está por trás do estudo das complexas redes de telefones de hoje.

Erlang era membro da Associação dos Matemáticos Dinamarqueses e foi através desta que estabeleceu contacto com outros matemáticos e com sócios da Companhia Telefónica de Copenhaga, para onde foi trabalhar em 1908, como colaborador científico e depois como chefe do seu laboratório.

Erlang começou a trabalhar aplicando a teoria de probabilidades a problemas de tráfego de telefones imediatamente e em 1909 surge o primeiro trabalho publicado sobre este: “The Theory of Probabilities and Telephone Conversations”, provando que ligações telefónicas distribuídas aleatoriamente seguiam a lei de distribuição de Poisson.

Mais tarde outros trabalhos foram publicados. O mais importante foi em 1917, “Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges”. Este continha fórmulas para perda e tempo de espera que agora são bem conhecidas na teoria de tráfego de telefones. Uma colecção dos principais trabalhos de Erlang nesta área pode ser encontrada na obra “The life and works of A.K. Erlang” [9], editada pela Companhia de Telefones de Copenhaga em 1948.

Segundo Syski [39], no início dos anos 50, as filas de espera eram geralmente estudadas com base em pressupostos Markovianos, em que o objectivo era obter uma expressão explícita para o comprimento da fila ou do tempo de espera, no estado de equilíbrio.

Em meados dos anos 60 alguns estudiosos, matemáticos e estatísticos, tinham alcançado resultados impressionantes apresentando soluções para praticamente todos os modelos de filas para um único servidor com entradas independentes, e resultados para sistemas de vários servidores, com tempo de serviço exponenciais. As filas de prioridade com chegadas com distribuição de Poisson e prioridades fixas foram também exaustivamente estudadas. E por tudo isto, em 1973, numa tese de Raph Disney é afirmado que não haveria mais nada para estudar na teoria das filas de espera (cf. [15]).

Realmente, nos anos 50 e 60 a teoria das filas de espera teve um estudo muito amplo e profundo. Entre muitos outros trabalhos destacam-se os trabalhos de Kendall [21], [22], Lindley [26], Smith [38], Cox [12] e Loynes [27], [28], discutido por Peter Whittle [43]. Outros vários textos foram publicados, como por exemplo os trabalhos de Morse [31], Syski [40], Saaty [37], Benes [7], Prabhu [34], Cohen [11], Gnedenko e Kovalenko [18]. Os dois volumes de Feller [16], [17], contêm muitos conhecimentos teóricos e aplicações para sistemas de filas, bem como em outras áreas aplicadas à probabilidade. Jaiswal [20] escreveu o primeiro livro totalmente dedicado às filas com prioridade. Em 1961 Cox e Smith [13] escreveram um livro notável resumindo a teoria das filas de espera existente.

A ideia de que a teoria das filas de espera já estava “completa”, caiu por terra, e ainda hoje existem muitos interessados, matemáticos, estatísticos a trabalhar em novos modelos da teoria das filas de espera, como o prova, por exemplo, a revista “Queueing Systems” (<http://www.springerlink.com/>).

O estudo das filas de espera tem aplicação imediata. Basta pensar que quando vamos a um supermercado, a um banco, a uma loja, ou se usarmos um simples elevador vamos encontrar um sistema de filas de espera. Designamos por sistema de fila uma estrutura que engloba a distribuição de chegada dos clientes ao sistema, a distribuição do tempo que um cliente demora para ser atendido dependendo do número de clientes que este encontra na fila de espera, a distribuição do serviço do cliente e o tempo que o cliente demora a ser servido e a sair do sistema. Obviamente, seria muito útil conhecer à partida, quando entramos no sistema o tempo que será necessário para estar servido, do ponto de vista do cliente. Do ponto de vista da entidade presta serviços, é importante conhecer qual o tamanho médio da fila e o tempo em que os servidores estão desocupados.

Nesta tese estudamos filas de espera com diferentes números de servidores, com condições diferentes relativamente à dimensão da população e à dimensão da “sala de espera”. Em cada caso estimamos o tempo médio que um cliente terá de esperar para ser atendido, sabendo quantos clientes se encontram no sistema e o tempo médio de cada serviço.

Este estudo resultou da exploração de algumas obras de probabilidade, de estatística, de processos estocásticos e, em particular, de filas de espera. O tema das filas de espera é, obviamente, muito vasto, e seria impossível querer resumir tudo que já existe sobre este tema numa só tese. Optámos então por apresentar os modelos markovianos e os não markovianos mais usuais. Passemos então à descrição resumida dos conteúdos dos restantes capítulos da tese.

No segundo capítulo expomos os conceitos e os resultados básicos que serão utilizados ao longo da tese.

No terceiro capítulo e seguintes, optámos por apresentar um texto que, embora incluindo, obviamente, definições e resultados, não os destaca, facilitando assim a sua leitura.

Introduzimos a teoria das filas de espera no terceiro capítulo, apresentando os elementos necessários para formar um modelo de filas. É ainda neste capítulo que fornecemos todas as ferramentas necessárias, providas da teoria dos processos de nascimento e morte, para trabalhar com as filas de espera. É também aqui que é introduzida a notação referente às principais variáveis aleatórias envolvidas nos capítulos seguintes. Damos a conhecer o que é o estado de equilíbrio e as famosas fórmulas de Little.

No capítulo seguinte exploramos com mais pormenor os modelos onde as chegadas de novos clientes ocorrem de acordo com um processo de Poisson, o serviço é uma variável aleatória com distribuição exponencial e a disciplina de atendimento é a FCFS (*First Come, First Served*), i.e., o primeiro cliente a chegar ao sistema é o primeiro a ser atendido e a sair do sistema. Estes modelos são, como veremos, os sistemas de mais simples “trato” e aqueles que, em muitas situações, se aproximam da realidade (recorde-se o destaque que a distribuição de Poisson tem como consequência da Lei dos Acontecimentos Raros e das suas propriedades como a estacionaridade, e a reguralidade). O que diferencia os vários modelos apresentados neste capítulo é o número de servidores (desde o caso em que o sistema tem um só servidor até o caso em que existem infinitos servidores), a capacidade do sistema e a capacidade da população. Quando a capacidade da população é infinita, esta não limita o

sistema. No entanto, podemos considerar sistemas onde a capacidade do sistema é infinita, mas a da população é finita e, neste caso, é a capacidade da população que limita o sistema. Na situação oposta temos os sistemas onde a capacidade do sistema é finita e a da população é infinita. Para cada um dos sistemas determinamos as características de cada usando as fórmulas de Little e elaboramos os respectivos gráficos.

No capítulo cinco estudamos outros modelos mais generalizados. Estudamos modelos onde as chegadas têm uma distribuição de Poisson e os serviços têm uma distribuição generalizada, não necessariamente exponencial. Estudamos também o modelo em que as chegadas têm uma distribuição generalizada (não seguindo necessariamente um processo de Poisson), e os serviços têm distribuição exponencial. E por fim estudamos modelos em que não são definidos nem a distribuição das chegadas nem dos serviços. Nestes modelos usando as fórmulas de Little conseguimos encontrar muitas propriedades destes, mas como é óbvio, não conseguimos ir tão longe como no capítulo anterior, e por vezes não conseguimos chegar a resultados exactos, mas só a aproximações.

2 Conceitos fundamentais

Para efectuar uma análise teórica de uma fila de espera teremos de ter em mente conceitos básicos e resultados da matemática, da estatística e da probabilidade que vamos utilizar. Estes conceitos e resultados podem ser encontrados em muitos livros da área, como por exemplo em [33].

2.1 Sigma-álgebra e função mensurável

Definição 1 Uma *experiência aleatória* é um procedimento obtido através de observações, onde é conhecido todo o conjunto dos possíveis resultados (Ω), qualquer realização de experiência produz um resultado (ω) que pertence a (Ω), e é incerto, pois não é possível sabê-lo sem fazer a experiência, porque existe uma intervenção do acaso.

Definição 2 O *espaço amostral* ou *espaço dos resultados* de uma experiência aleatória, Ω , é o conjunto de todos os acontecimentos resultados concebíveis de uma experiência aleatória.

Definição 3 Dados um conjunto Ω e um subconjunto do conjunto das partes de Ω , $\mathcal{A} \subseteq P(\Omega)$, dizemos que \mathcal{A} é uma *sigma-álgebra* sobre Ω se

1. $\Omega \in \mathcal{A}$;
2. Se $E \in \mathcal{A}$, então $\overline{E} = \Omega \setminus E \in \mathcal{A}$;
3. Se, $\forall n \in \mathbb{N}$, $E_n \in \mathcal{A}$, então $\bigcup_{n \in \mathbb{N}} E_n \in \mathcal{A}$.

O par (Ω, \mathcal{A}) é designado por *espaço mensurável* e os elementos de \mathcal{A} conjuntos mensuráveis relativamente à sigma-álgebra \mathcal{A} .

A sigma-álgebra usual sobre \mathbb{R} é a sigma-álgebra dos **borelianos**, \mathcal{B} , isto é, a sigma-álgebra gerada pela topologia usual de \mathbb{R} ou, equivalentemente, gerada pela classe $\{(-\infty, a], a \in \mathbb{R}\}$. Por outras palavras \mathcal{B} é a menor sigma-álgebra que contém a topologia τ_μ .

Definição 4 Os elementos de uma sigma-álgebra são designados por *acontecimentos*.

Definição 5 Seja A um acontecimento de um espaço mensurável (Ω, \mathcal{A}) . O *acontecimento complementar* de A é $\Omega - A = \Omega \setminus A = \overline{A} = A^C$.

Definição 6 Dados dois espaços mensuráveis $(\Omega_1, \mathcal{A}_1)$ e $(\Omega_2, \mathcal{A}_2)$ e f uma aplicação definida de Ω_1 para Ω_2 , dizemos que f é uma **função mensurável** se a pré-imagem de qualquer conjunto mensurável (relativamente a \mathcal{A}_2) for um conjunto mensurável (relativamente a \mathcal{A}_1) isto é, se

$$\forall B \in \mathcal{A}_2, f^{-1}(B) \in \mathcal{A}_1.$$

2.2 Independência de acontecimentos

Definição 7 Consideremos uma aplicação μ não negativa, definida numa sigma-álgebra \mathcal{A} , sobre um conjunto Ω . Dizemos que

$$\mu : \mathcal{A} \longrightarrow [0, +\infty]$$

é uma **medida** se:

1. $\mu(\emptyset) = 0$;
2. Dada uma sucessão $\{E_i\}_{i \in \mathbb{N}}$ de conjuntos mensuráveis disjuntos, dois a dois, temos

$$\mu\left(\bigcup_{n \in \mathbb{N}} E_n\right) = \sum_{n \in \mathbb{N}} \mu(E_n).$$

Definição 8 Se μ for uma medida definida num espaço mensurável (Ω, \mathcal{A}) e se $\mu(\Omega) = 1$ dizemos que μ é uma **medida de probabilidade**.

As medidas de probabilidade denotam-se usualmente por P .

O terno (Ω, \mathcal{A}, P) é designado por espaço de medida. Assumiremos no resto deste capítulo estar a trabalhar num espaço de medida representado por (Ω, \mathcal{A}, P) .

Definição 9 Seja B um acontecimento tal que $P(B) > 0$. Definimos a probabilidade de A sabendo (ou **condicionado a**) B da seguinte forma:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Definição 10 Sejam A e B dois acontecimentos. Dizemos que A e B são **independentes** se

$$P(A \cap B) = P(A)P(B).$$

Neste caso a ocorrência de B não tem qualquer efeito sobre a probabilidade de acontecer A e vice-versa. De facto, se A e B são independentes então

$$\begin{aligned} P(A|B) &= P(A), \text{ se } P(B) > 0 \\ &\text{e} \\ P(B|A) &= P(B), \text{ se } P(A) > 0. \end{aligned}$$

Mais geralmente dizemos que os acontecimentos A_1, A_2, \dots, A_n ($n \in \mathbb{N}_1$) são independentes se para qualquer $k \leq n$, para cada escolha de k acontecimentos, a probabilidade da sua intersecção é igual ao produto das suas probabilidades.

Teorema 11 (da Probabilidade Total) Seja $\{B_i\}_{i \in \mathbb{N}}$ uma partição do espaço de resultados Ω , com $P(B_i) > 0, \forall i \in \mathbb{N}$. Dado um qualquer acontecimento A , temos que:

$$P(A) = \sum_{i \in \mathbb{N}} P(A|B_i) P(B_i). \quad (1)$$

2.3 Variável Aleatória

Definição 12 Dados um espaço amostral Ω e uma sigma-álgebra \mathcal{A} sobre este, dizemos que $X : \Omega \longrightarrow \mathbb{R}$ é uma **variável aleatória** (v.a.) se X for mensurável relativamente a \mathcal{A} e a \mathcal{B} , a sigma-álgebra dos borelianos.

Definição 13 Seja X uma v.a., definida num espaço mensurável (Ω, \mathcal{A}) munida de uma probabilidade P . Chamamos função de distribuição acumulada (f.d.a.), ou simplesmente **função de distribuição** (f.d.) de X à função real de variável real definida por

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto P[X \leq x] := P[X^{-1}((-\infty, x])] \end{aligned} .$$

Dizemos que X e Y , duas variáveis aleatórias, são *identicamente distribuídas* e escrevemos $X \stackrel{d}{=} Y$, se $F_X = F_Y$.

Definição 14 Seja X uma v.a. Dizemos que X é uma **v.a. discreta** se o conjunto dos valores que toma é discreto. Neste caso X pode ser representada da seguinte forma:

$$X = \begin{cases} x_k, & k \in K & \leftarrow \text{valores que } X \text{ toma} \\ P_k = P[X = x_k] & & \leftarrow \text{probabilidade de a v.a. } X \text{ tomar esses valores} \end{cases} .$$

A função anterior é a função massa de probabilidade (f.m.p.) de X .

Uma função $x_k \rightarrow f(x_k)$ ($k \in K$) é uma **função massa de probabilidade** (f.m.p.) se

$$\begin{cases} \sum_{k \in K} f(x_k) = 1 \\ 0 \leq f(x_k) \leq 1, \forall k \in K \end{cases} .$$

Definição 15 Dizemos que X é uma **v.a. contínua**, se F_X for contínua.

Definição 16 Dizemos que X é uma **v.a. absolutamente contínua** se existir uma função não negativa $f_X : \mathbb{R} \longrightarrow \mathbb{R}$ tal que, $\forall x \in \mathbb{R}$,

$$F_X(x) = P[X \leq x] = \int_{-\infty}^x f_X(u) du.$$

A função f_X é chamada uma **função densidade de probabilidade** (f.d.p.) de X .

Dada uma função f , f será uma f.d.p. de alguma v.a. se

$$\begin{cases} f_X \geq 0 \\ \int_{\mathbb{R}} f_X(x) dx = 1 \end{cases}.$$

Definição 17 Dizemos que X_1, X_2, \dots, X_n são v.a.'s **independentes** se os acontecimentos $X_1 \in A_1, \dots, X_n \in A_n$ são independentes, $\forall n \in \mathbb{N}_1, \forall A_1, \dots, A_n \in \mathcal{B}$.

Quando, além de independentes, as v.a.'s X_1, X_2, \dots, X_n têm a mesma distribuição dizemos que são independentes e identicamente distribuídas (iid).

Definição 18 Dada uma v.a. X , o **valor esperado, valor médio, ou a esperança** de X é definido por

$$E[X] = \int_{\mathbb{R}} x dP_X(x), \text{ se } \int_{\mathbb{R}} |x| dP_X(x) < +\infty,$$

onde P_X é medida induzida por X , isto é, a medida que a cada conjunto boreliano B atribui o valor $P[X^{-1}(B)]$.

Se $\int_{\mathbb{R}} |x| dP_X(x) = +\infty$, não existe valor esperado.

- No caso de a v.a. X ser uma v.a. discreta, o valor esperado é dado por

$$E[X] = \sum_{k \in K} x_k P[X = x_k], \text{ se } \sum_{k \in K} |x_k| P[X = x_k] < +\infty.$$

Se $\sum_{k \in K} |x_k| P[X = x_k] = +\infty$, não existe valor esperado.

- No caso de a v.a. X ser uma v.a. absolutamente contínua com f.d.p. f_X , temos

$$E[X] = \int_{\mathbb{R}} x f_X(x) dx, \text{ se } \int_{\mathbb{R}} |x| f_X(x) dx < +\infty.$$

Se $\int_{\mathbb{R}} x f_X(x) dx = +\infty$, não existe valor esperado.

Mais geralmente,

Definição 19 Dada uma v.a. X , o **momento de ordem n** (ou o **n -ésimo momento**) de X , $n \in \mathbb{N}$, é definido por

$$E[X^n] = \int_{\mathbb{R}} x^n dP_X(x), \text{ se } \int_{\mathbb{R}} |x|^n dP_X(x) < +\infty,$$

onde P_X é medida induzida por X .

Se $\int_{\mathbb{R}} |x|^n dP_X(x) = +\infty$, não existe o momento de ordem n .

Mais geralmente, se $g : \mathbb{R} \rightarrow \mathbb{R}$ for uma função mensurável

$$E[g(x)] = \int_{\mathbb{R}} g(x) dP_X(x), \text{ se } \int_{\mathbb{R}} |g(x)| dP_X(x) < +\infty.$$

Neste trabalho surgirão dois valores esperados especiais: a função geradora de probabilidades e a transformada de Laplace.

Definição 20 A **função geradora de probabilidades** (f.g.p.) de uma v.a. X discreta é dada por

$$P(z) = E[z^X] = \sum_{n=0}^{\infty} P[X=n] z^n = \sum_{n=0}^{\infty} P_n z^n,$$

com $|z| \leq 1$.

Uma propriedade da f.g.p é

$$\left. \frac{d^{(n)}}{dz^n} P(z) \right|_{z=0} = n! P[X=n] \Rightarrow P[X=n] = \frac{1}{n!} \left. \frac{d^{(n)}}{dz^n} P(z) \right|_{z=0}.$$

Definição 21 Chamamos **transformada de Laplace** da função de distribuição F_X , ou mais geralmente, da v.a. X , onde $X > 0$, a

$$L_X(t) = E[e^{-tX}] = \int_0^{\infty} e^{-tx} dF_X(x).$$

Definição 22 Seja X uma v.a.. Definimos a **variância** de X da forma seguinte, caso exista,

$$\text{var}[X] = E[(X - E[X])^2].$$

- No caso de a v.a. X ser uma v.a. discreta,

$$\text{var}[X] = \sum_{k \in K} (x_k - E[X])^2 P[X = x_k].$$

- No caso de a v.a. X ser uma v.a. absolutamente contínua,

$$\text{var}[X] = \int_{\mathbb{R}} (x - E[X])^2 f_X(x) dx.$$

Não é difícil mostrar que

$$\text{var}[X] = E[X^2] - E^2[X].$$

Definição 23 Sejam X e Y duas v.a.'s.. Chamamos **covariância** entre X e Y ao valor definido da seguinte forma:

$$\text{cov}[X, Y] = E[(X - E[X]) \cdot (Y - E[Y])],$$

(supondo que existem os valores esperados envolvidos na expressão).

Referimos seguidamente quatro distribuições que terão algum destaque ao longo deste trabalho:

Definição 24 Dizemos que X tem **distribuição de Poisson** com parâmetro $\lambda > 0$, ($X \sim \text{Poisson}(\lambda)$) se

$$P[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, \dots$$

Não é difícil mostrar que

$$E[X] = \sum_{k=0}^{+\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda,$$

e que

$$\text{var}[X] = E[X].$$

Definição 25 Uma v.a. X tem **distribuição exponencial** com parâmetro $\alpha > 0$, ($X \sim \text{Exponencial}(\alpha)$) se tem função densidade de probabilidade definida por:

$$f(t) = \begin{cases} \alpha e^{-\alpha t}, & \text{se } t > 0 \\ 0, & \text{se } t \leq 0 \end{cases}.$$

A distribuição acumulada, é

$$F(t) = P([T \leq t]) = \begin{cases} 1 - e^{-\alpha t}, & \text{se } t > 0 \\ 0, & \text{se } t \leq 0 \end{cases}.$$

Facilmente se mostra que:

$$E[X] = \frac{1}{\alpha}$$

e que

$$\text{var}[X] = \frac{1}{\alpha^2},$$

donde resulta que, $E[X] = \sqrt{\text{var}(X)}$.

É habitual dizer-se que a distribuição exponencial não tem memória, pois

$$P[X > u + t | X > u] = P[X > t], \quad (2)$$

com $u, t > 0$.

Definição 26 Uma v.a. X tem **distribuição de Erlang** com parâmetros $r \in \mathbb{N}_1$ e $\lambda > 0$, ($X \sim \text{Erlang}(r, \lambda)$), se a função densidade é dada por, para $t > 0$,

$$f(t) = \frac{\lambda^r t^{r-1} e^{-\lambda t}}{(r-1)!}$$

Neste caso temos que

$$E[X] = \frac{r}{\lambda}$$

e que

$$\text{var}[X] = \frac{r}{\lambda^2}.$$

Mais geralmente,

Definição 27 Uma v.a. X tem **distribuição gama** com parâmetros α e β , ($X \sim \text{Gama}(\alpha, \beta)$), $\alpha, \beta > 0$ se a sua função densidade é dada por

$$f(t) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}, & \text{se } t > 0 \\ 0, & \text{se } t \leq 0 \end{cases},$$

onde Γ é a função gama definida, para $\alpha > 0$, por,

$$\Gamma(\alpha) = \int_0^{+\infty} e^{-t} t^{\alpha-1} dt$$

(para consultar algumas propriedades da função Γ consultar, por exemplo [33]). Facilmente se mostra que

$$E[X] = \frac{\alpha}{\beta}$$

e que

$$\text{var}[X] = \frac{\alpha}{\beta^2}.$$

Note-se que quando $\alpha = 1$, $X \sim \text{Exponencial}(\beta)$ e que quando $\alpha \in \mathbb{N}_1$, $X \sim \text{Erlang}(\alpha, \beta)$.

Entre a distribuição exponencial e a distribuição gama (Erlang) existe uma outra ligação que será utilizada neste trabalho: dadas n v.a.'s X_1, X_2, \dots, X_n , independentes e idênticas com distribuição comum $\text{Exponencial}(\lambda)$, a sua soma tem distribuição $\text{Gama}(n, \lambda) \equiv \text{Erlang}(n, \lambda)$, isto é,

$$\sum_{i=1}^n X_i \sim \text{Gama}(n, \lambda).$$

A prova deste resultado e dos resultados acima referidos podem ser encontrados por exemplo, em [33].

2.4 Processos estocásticos

Definição 28 Um **processo estocástico** é um conjunto de variáveis aleatórias X_t ou $X(t)$, onde t é um parâmetro que percorre um conjunto de índices T . Geralmente, toma-se para conjunto T um dos conjuntos: \mathbb{N}_0 ou $[0, +\infty)$, e neste último caso, a situação mais comum será que t represente o tempo.

Os processos estocásticos são distinguidos pelo seu **espaço de estados**, ou seja, pelo conjunto dos valores que as v.a.'s envolvidas X_t podem tomar, pelo seu conjunto de índices T e pelas relações de dependência entre as v.a.'s X_t , $t \in T$.

Um processo estocástico que vamos usar ao longo do trabalho e que tem inúmeras aplicações é o processo de Markov.

Definição 29 Um **processo de Markov** $\{X_t\}_{t \in T}$ é um processo estocástico onde a probabilidade de qualquer comportamento futuro do processo, quando o estado actual é conhecido, não é alterado pelo conhecimento adicional relativo ao comportamento passado. Isto é, para os instantes arbitrários, $t_1 < t_2 < \dots < t_k < t_{k+1}$:

$$P[X_{t_{k+1}} = x_{k+1} | X_{t_k} = x_k, \dots, X_{t_1} = x_1] = P[X_{t_{k+1}} = x_{k+1} | X_{t_k} = x_k], \quad (3)$$

se X_t for uma v.a. discreta;

$$P[a \leq X_{t_{k+1}} \leq b | X_{t_k} = x_k, \dots, X_{t_1} = x_1] = P[a \leq X_{t_{k+1}} \leq b | X_{t_k} = x_k], \quad (4)$$

se X_t for absolutamente contínua.

Um processo de Markov particular é o processo de Poisson:

Definição 30 Um **processo de Poisson** com intensidade, taxa, ou número médio de acontecimentos por unidade de tempo, $\lambda > 0$, é um processo estocástico $\{X_t\}_{t \geq 0}$ que toma valores inteiros não negativos tal que

- $X_0 = 0$;
- Para $s \geq 0$ e $t > 0$, a variável aleatória $X_{s+t} - X_s$ tem distribuição de Poisson com parâmetro λt :

$$P(X_{s+t} - X_s = k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}, \text{ para } k = 0, 1, \dots$$

onde t é o comprimento do intervalo de tempo em análise e k é o número de acontecimentos que ocorrem no intervalo t ;

- Para qualquer colecção de instantes $t_0 = 0 < t_1 < t_2 < \dots < t_n$, os incrementos do processo

$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}},$$

são variáveis aleatórias independentes.

Da definição resulta que se $\{X_t\}_{t \geq 0}$ é um processo de Poisson de taxa $\lambda > 0$, então a média e a variância de X_t é

$$E[X_t] = \text{var}(X_t) = \lambda t.$$

Como referimos, um processo de Poisson é um processo de Markov pois satisfaz a condição (4):

$$\begin{aligned} & P[X_{t_{k+1}} = j | X_{t_k} = i, X_{t_{k-1}} = x_{k-1}, \dots, X_{t_1} = x_1] \\ &= P[j - i \text{ eventos em } (t_k, t_{k+1}]] \\ &= P[X_{t_{k+1}} = j | X_{t_k} = i]. \end{aligned}$$

O processo de Poisson é também um processo homogéneo, pois as probabilidades de transição $p_{ij}(s) = P(X_{t+s} = j | X_t = i)$ são independentes do instante t ($t, s > 0$).

O tempo O_1 até a ocorrência do primeiro acontecimento tem distribuição dada por

$$\begin{aligned} F_{O_1}(t) &= P[O_1 \leq t] \\ &= 1 - P[O_1 > t] \\ &= \begin{cases} 1 - P[X_t = 0], & \text{se } t > 0 \\ 1 - 1, & \text{se } t \leq 0 \end{cases} \\ &= \begin{cases} 1 - e^{-\lambda t}, & \text{se } t > 0 \\ 0, & \text{se } t \leq 0 \end{cases}, \end{aligned}$$

isto é, $O_1 \sim \text{Exponencial}(\lambda)$.

Mas se O_r for o tempo que decorre entre a realização do $(r-1)$ -ésimo acontecimento e do r -ésimo acontecimento temos, pelo mesmo raciocínio, $O_r \sim \text{Exponencial}(\lambda)$, de onde se conclui que $O_r \stackrel{d}{=} O$, $\forall r \in \mathbb{N}_1$, onde $O \sim \text{Exponencial}(\lambda)$.

Tendo em conta este facto e a independência entre as variáveis aleatórias O_1, O_2, \dots resulta que o tempo decorre até a ocorrência do r -ésimo acontecimento tem distribuição de *Erlang* (r, λ) , $r \in \mathbb{N}_1$, $\lambda > 0$.

2.4.1 Processo de vida e morte

Processo de nascimento

Consideremos uma sucessão de números positivos $\{\lambda_i\}_{i \in \mathbb{N}}$. Definimos um processo de nascimento puro como sendo um processo de Markov satisfazendo as seguintes condições ($k = 0, 1, 2, \dots$):

1. $P_{k,k+1}(h) = P[X(t+h) = k+1 | X(t) = k] = \lambda_k h + o(h)$, para $h \downarrow 0$;
2. $P_{k,k}(h) = P[X(t+h) = k | X(t) = k] = 1 - \lambda_k h + o(h)$, para $h \downarrow 0$;
3. $P[X(t+h) < k | X(t) = k] = 0$, para $h > 0$;
E, por conveniência, muitas vezes consideramos que
4. $X(0) = 0$.

Recorde-se que $o(h)$ denota uma função f tal que

$$\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0.$$

Tomemos $P_n(t) = P[X(t) = n]$ e suponhamos que $X(0) = 0$. Podemos escrever $P_n(t)$ como função dos parâmetros do modelo $(\lambda_0, \lambda_1, \lambda_2, \dots)$.

A solução depende de os parâmetros do modelo serem todos diferentes ou não. Mostra-se que, (ver, por exemplo, [42]) no caso em que os parâmetros são diferentes,

$$\begin{aligned} P_n(t) &= P[X(t) = n | X(0) = 0] \\ &= \lambda_0 \dots \lambda_{n-1} [B_{0,n} \exp(-\lambda_0 t) + \dots + B_{n,n} \exp(-\lambda_n t)] \\ &= \prod_{i=0}^{n-1} \lambda_i \sum_{i=0}^n [B_{i,n} \exp(-\lambda_i t)], \quad n \geq 1 \end{aligned}$$

onde

$$B_{k,n} = \frac{1}{\prod_{\substack{j=0 \\ j \neq k}}^n (\lambda_j - \lambda_k)}.$$

Processo de morte

Complementando o processo de nascimento existe o processo de morte. Este move-se sucessivamente pelos estados $N, N-1, \dots, 2, 1$ e acabará por ser absorvido no estado 0 (estado correspondente à extinção da população). O processo é especificado pelos parâmetros de morte $\mu_k > 0$, para $k = 1, 2, \dots, N$, onde os tempos entre o k e o $(k+1)$ -ésimos eventos são distribuídos exponencialmente com parâmetro μ_k , sendo estes tempos independentes.

Alternativamente, podemos descrever um processo de morte como sendo um processo de Markov $\{X_t\}_{t \geq 0}$ cujos estados são 0, 1, 2, ..., N e para o qual

1. $P_{k,k-1}(h) = P[X(t+h) = k-1 | X(t) = k] = \mu_k h + o(h)$, para $h \downarrow 0$, $k = 1, 2, \dots, N$;
2. $P_{k,k}(h) = P[X(t+h) = k | X(t) = k] = 1 - \mu_k h + o(h)$, para $h \downarrow 0$, $k = 1, 2, \dots, N$;
3. $P[X(t+h) > k | X(t) = k] = 0$, $k = 0, 1, \dots, N$.

Por convenção toma-se $\mu_0 = 0$. O parâmetro μ_k é a taxa de morte quando o processo está no estado k . Quando os parâmetros de morte $\mu_1, \mu_2, \dots, \mu_N$ são distintos, temos

$$P_N(t) = \exp(-\mu_N t)$$

e, para $n < N$,

$$\begin{aligned} P_n(t) &= P[X(t) = n | X(0) = N] \\ &= \mu_{n+1} \mu_{n+2} \dots \mu_N [A_{n,n} \exp(-\mu_n t) + \dots + A_{N,n} \exp(-\mu_N t)] \\ &= \prod_{i=n+1}^N \mu_i \sum_{i=n}^N [A_{i,n} \exp(-\mu_i t)], \end{aligned}$$

onde

$$A_{k,n} = \frac{1}{\prod_{\substack{j=n \\ j \neq k}}^N (\mu_j - \mu_k)}.$$

Processo de nascimento e morte

Uma generalização óbvia dos processos de nascimento e dos processos de morte é obtida quando permitimos que o processo $\{X(t)\}_{t \geq 0}$ possa tanto diminuir como aumentar. Logo, se num instante t o processo se encontra no estado n , poderá, depois de um tempo aleatório, mover-se para um dos seus estados vizinhos $n + 1$ ou $n - 1$.

Mais especificamente, se $\{X(t)\}_{t \geq 0}$ é um processo de nascimento e morte com parâmetro de nascimento λ_k e parâmetro de morte μ_k temos:

1. $P_{k,k+1}(h) = \lambda_k h + o(h)$, para $h \downarrow 0$, $k \geq 0$;
2. $P_{k,k-1}(h) = \mu_k h + o(h)$, para $h \downarrow 0$, $k \geq 1$;
3. $P_{k,k}(h) = 1 - (\lambda_k + \mu_k) h + o(h)$, para $h \downarrow 0$, $k \geq 1$;
4. $P_{i,j}(0) = \delta_{ij}$;
5. $\mu_0 = 0$, $\lambda_0 > 0$, $\mu_k, \lambda_k > 0$, $k \in \mathbb{N}_1$.

O processo de vida e morte é um processo estocástico “sem memória” baseado na distribuição exponencial que associa, no contexto das filas de espera, uma **vida** a uma chegada à fila e uma **morte** à saída de um cliente depois de ser atendido.

Distribuição estacionária do processo de nascimento e morte

Para um processo geral de nascimento e morte que não tem estados absorventes (quando o processo permanece neste depois de o alcançar, com probabilidade 1), é possível mostrar que os limites

$$\lim_{t \rightarrow +\infty} p_{ij}(t) = P_j \geq 0, \text{ com } p_{ij} \text{ é a probabilidade de transição,} \\ \text{do estado } i \text{ para o estado } j$$

existem e são independentes do estado inicial i (ver [42]). Pode acontecer que $P_j = 0$, para todos os estados j . Quando os limites $P_j > 0$, e satisfazem $\sum_{j \geq 0} P_j = 1$, estes formam uma distribuição de probabilidade que é, naturalmente, chamada de distribuição limite do processo. A distribuição limite é estacionária no sentido em que

$$P_j = \sum_{i=0}^{\infty} P_i p_{ij}(t).$$

Parte da importância dos processos de nascimento e de morte como modelos advém da existência de fórmulas que nos permitem determinar se existe ou não distribuição limite e, quando existe, que nos permitem também encontrar a distribuição limite. No processo de nascimento e morte apresentado na secção anterior, a distribuição estacionária é encontrada através das igualdades

$$\begin{aligned} 0 &= -\lambda_0 P_0 + \mu_1 P_1, \\ 0 &= \lambda_{j-1} P_{j-1} - (\lambda_j + \mu_j) P_j + P_{j+1} P_{j+1}, \quad j \geq 1. \end{aligned} \tag{5}$$

A solução de (5) é obtida por indução. Tomando $\theta_0 = 1$ e

$$\theta_j = \frac{\lambda_0 \lambda_1 \dots \lambda_{j-1}}{\mu_1 \mu_2 \dots \mu_j}$$

para $j \geq 1$, temos que $P_j = \theta_j P_0$. Para que $\{P_j\}_{j \in \mathbf{N}}$ defina uma função massa de probabilidade, é necessário que

$$\sum_{j \geq 0} P_j = \sum_{j \geq 0} \theta_j P_0 = P_0 \sum_{j \geq 0} \theta_j = 1 \Rightarrow P_0 = \frac{1}{\sum_{j \geq 0} \theta_j}.$$

Se $\sum_{j \geq 0} \theta_j < \infty$, obtemos então

$$P_j = \frac{\theta_j}{\sum_{j \geq 0} \theta_j}.$$

Se $\sum_{j \geq 0} \theta_j = \infty$, obtemos $P_j = 0$, para todo o valor j , o que significa que não existe distribuição limite.

3 Teoria das filas de espera

As filas de espera são um fenómeno corrente no dia-a-dia, onde existem clientes que desejam prestações de serviços (servidores que para serem utilizados é necessário que os clientes esperem e formem uma fila física ou conceptual) e que haja um número de clientes superior ao número de servidores, pois o servidor demora algum tempo a atender cada cliente (tempo de serviço), e este serviço termina quando o cliente se retira.

Os pontos de interesse da teoria das filas são: o tempo de espera do cliente, o número de clientes na fila e a razão entre o tempo de espera e o tempo de prestação de serviço.

Existem diversas aplicações da teoria das filas. Entre elas destacam-se: o fluxo de tráfego (aviões, carros, pessoas, comunicações,...) e as prestação de serviços (bancos, correios,...).

A teoria das filas de espera tem como objectivo “optimizar” o funcionamento das filas de espera, encontrando soluções equilibradas entre dois extremos: o *congestionamento* que acontece quando os clientes têm que esperar demasiado tempo na fila que só é aceitável quando o custo do servidor é muito maior do que o custo de espera do cliente e a *rarefacção* que acontece quando os servidores permanecem inactivos durante por muito tempo, que pode ser uma situação desejável (como por exemplo nos serviços de bombeiros) ou indesejável (como por exemplo num supermercado).

3.1 Elementos constituintes de um sistema

Para formar um sistema de fila de espera existem cinco elementos fundamentais que podem tomar vários formatos: a população, a fila de espera, o serviço, a capacidade do sistema e a disciplina do atendimento.

3.1.1 Chegadas, Fonte ou População

As chegadas, fonte ou população é o elemento que gera os clientes que vão chegar ao sistema, que pode ter características muito diferentes, como seguidamente se descreve.

Dimensão da população A dimensão da população pode ser finita ou infinita. Neste último caso a probabilidade de ocorrer uma nova chegada não é influenciada pelo número de clientes que já se encontram no sistema.

Dimensão da chegada A dimensão da chegada pode ser unitária quando os clientes chegam um a um ou em grupo.

Controlo das chegadas O processo das chegadas pode ser controlável (como por exemplo quando existem inscrições em dias fixos) ou incontrolável (como por exemplo numa urgência de um hospital).

Distribuição das chegadas A distribuição das chegadas pode ser descrita pelo tempo de entre duas chegadas seguidas ou pela quantidade de chegadas por unidade de tempo. Podem ser constantes quando existem intervalos de tempo fixos entre chegadas sucessivas (como por exemplo em filas de montagem industrial) ou aleatórias quando os intervalos de tempo entre chegadas sucessivas não podem ser previstos, usando-se neste caso distribuições de probabilidade.

Taxa de chegada A taxa de chegada é o número médio de clientes que procuram o serviço por unidade de tempo. Esta taxa habitualmente denotada por λ pode ser independente do número de clientes existente no sistema ou dependente deste. Neste último caso, se o número de clientes for, por exemplo, n , escrevemos λ_n .

Atitude dos clientes A atitude dos clientes pode ser paciente, quando estes permanecem na fila até serem atendidos, ou impaciente quando desistem de esperar e saem do meio da fila ou então nem se juntam à fila se esta estiver muito grande.

3.1.2 Fila de espera

As características da fila de espera também podem variar de sistema para sistema, como mostramos a seguir.

Número de filas O número de filas num sistema pode variar entre uma fila simples, com uma única fila mesmo que o servidor tenha vários postos de atendimento, ou fila múltipla que consiste em uma fila para cada posto de atendimento.

Comprimento da fila O comprimento da fila pode ser finito ou infinito.

Disciplina da fila A disciplina da fila pode ser variada. A mais comum é o primeiro cliente a entrar na fila começar a formar a fila e os seguintes ficarem de seguida a este. Pode também ser uma fila com prioridade para reservas e para emergências, necessidades limitadas ou outras.

3.1.3 Serviço

As características do serviço e do tipo de serviço também são importantes e podem variar:

Configuração de serviço A configuração de serviço corresponde ao número de servidores em paralelo (postos de atendimento), e do número de fases de atendimento.

Dimensão de serviço A dimensão de serviço pode ser simples (como por exemplo num banco) ou em grupo quando vários clientes podem ser atendidos ao mesmo tempo no mesmo servidor (como por exemplo acontece num elevador).

Distribuição do tempo de serviço O tempo de serviço pode ser constante ou aleatório. As distribuições aleatórias mais estudadas são a exponencial e a Erlang.

Taxa de serviço A taxa de serviço corresponde ao número médio de clientes que podem ser atendidos por cada servidor e por unidade de tempo e é habitualmente denotada por μ . Esta taxa pode ser independente do número de clientes existente no sistema ou dependente do número de clientes e, nesse caso, se o número de clientes for, por exemplo n , escrevemos μ_n .

3.1.4 Capacidade do sistema

A capacidade do sistema corresponde ao número máximo de clientes que o sistema suporta, incluindo os que estão em espera e os que estão sendo atendidos. Pode ser infinita ou finita. Caso seja finita e o sistema esteja cheio, não pode entrar nenhum cliente antes que outro saia.

3.1.5 Disciplina de atendimento

A disciplina de atendimento descreve a forma como os clientes saem da fila de espera para serem atendidos. Existem várias formas:

FCFS (*First Come, First Served*) ou FIFO (*First In, First Out*) As filas com características FCFS (ou FIFO) são as filas onde o primeiro cliente a chegar é o primeiro a ser atendido e a sair. Estas são as filas mais comuns na vida diária (como por exemplo acontece com as filas num banco).

LCFS (*Last Come, First Served*) ou LIFO (*Last In, First Out*) As filas LCFS (ou LIFO) são as filas onde o último cliente a chegar é o primeiro a ser atendido e a sair (como por exemplo acontece nas pilhas da Torre de Hanoi).

SIRO (*Service In Random Order*) As filas SIRO são filas em que o serviço é feito de forma aleatória.

PRI (Prioritárias) As filas PRI são as filas com prioridade, onde é atribuída uma prioridade a cada cliente, podendo um cliente que entra com maior prioridade ser atendido imediatamente, interrompendo o atendimento do cliente que está a ser atendido nesse momento (como por exemplo pode acontecer num serviço de emergência médica) ou o cliente com maior prioridade ser colocado no início da fila, e sendo o próximo a ser atendido após a saída do cliente que está nesse momento a ser atendido (como por exemplo é a prioridade dada às grávidas em certos serviços).

RR (*Round-Robin*) Nas filas que seguem a regra de Round-Robin cada cliente recebe uma fatia de tempo do serviço, dentro da qual é atendido. Após terminar esse tempo, mesmo que a actividade não tenha sido completada, o cliente é retirado e outro passa a ser atendido. O cliente cujo serviço foi interrompido, retorna ao serviço, posteriormente (como por exemplo num lugar que só tem um computador e os clientes só têm direito a 30 minutos de cada vez).

GD (*General Discipline*) As filas GD seguem uma disciplina genérica, ou seja, nestas filas não é especificada a disciplina de atendimento.

3.2 Tipos de sistemas de filas de espera

Existem muitos tipos de sistemas de filas de espera, pois basta fazer várias combinações dos elementos referidos anteriormente para formar uma fila de espera diferente.

3.2.1 Variantes das filas de espera

Segundo Müller *et al.* [32] em 1953, o matemático inglês David George Kendall (1918 – 2007) propôs a seguinte notação para representar cada fila de espera:

$$A/S/m/K/N/Q$$

onde:

- A é a distribuição dos tempos entre as chegadas (processo de chegadas);
- S é a distribuição dos tempos de serviço (processo de atendimento);
- m é o número de servidores ($m \in \mathbb{N}$);
- K é a capacidade do sistema ($K \in \mathbb{N} \cup \{+\infty\}$);
- N é o tamanho da população ($N \in \mathbb{N} \cup \{+\infty\}$);
- Q é a disciplina de atendimento.

Muitas vezes, os três últimos símbolos são omitidos. Nestes casos, assume-se capacidade ilimitada, população infinita e disciplina de atendimento FCFS.

Ao longo deste trabalho usaremos as seguintes letras para as distribuições indicadas: M (de Memoryless) para a distribuição exponencial, estejamos a falar da distribuição dos tempos entre chegadas de clientes ou da distribuição do tempo de serviço; E_r (de Erlang) para a distribuição de Erlang e G (de Genérica) para uma distribuição genérica.

3.3 Características das filas de espera

Em todas as filas de espera existem características muito úteis, das quais citamos algumas, indicando a sua notação habitual. Em particular, a letra L será usada em questões relacionadas com o comprimento da fila (L da palavra inglesa *Length* - Comprimento) e a letra W será usada em questões relacionadas com tempo de espera (W da palavra inglesa *Waiting* - esperando). Da mesma forma usamos as letras q e s para indicar a fila (*queue* em inglês) e o sistema (*system* em inglês), respectivamente.

- Taxa de chegada (λ), isto é, número médio de clientes que chegam por unidade de tempo. Como referimos anteriormente, quando a taxa de chegadas depende do número de clientes n , designamo-la por λ_n ;
- Intervalo médio entre duas chegadas consecutivas ($\frac{1}{\lambda}$);
- Taxa média de entrada no sistema ($\lambda' = \sum_{n=0}^{\infty} \lambda_n P_n$), com P_n a probabilidade de existirem n clientes no sistema a longo prazo, isto é, quando o sistema está em equilíbrio;
- Taxa de serviço (μ), número médio de clientes que cada servidor atende por unidade de tempo. Como referimos anteriormente, quando a taxa de serviço depende do número de clientes n , designamo-la por μ_n ;
- Tempo médio de serviço ($\frac{1}{\mu}$);
- Comprimento médio da fila (L_q) (sem incluir os clientes que estão a ser atendidos);
- Número médio de clientes no sistema (L_s);
- Tempo médio de espera na fila (W_q) (este tempo exclui o tempo que o cliente leva a ser atendido);
- Tempo médio de espera no sistema (W_s);
- Taxa média de ocupação do serviço (ρ), percentagem de tempo durante o qual o serviço está ocupado;
- Taxa média de desocupação do serviço ($1 - \rho$), percentagem de tempo durante o qual o serviço está desocupado;
- Número de servidores (S);
- Probabilidade de existirem no sistema k ou mais clientes ($P[N \geq k] = \sum_{n=k}^{\infty} P_n$);
- Probabilidade de o tempo de espera na fila ser zero ($P[T_q = 0]$);
- Probabilidade de o tempo de espera na fila exceder t ($P[T_q > t]$);
- Probabilidade de o tempo gasto no sistema exceder t ($P[T_s > t]$).

3.3.1 Estado estacionário ou de equilíbrio

Para que ocorra uma situação estacionária é preciso que o número médio de entradas por unidade de tempo seja igual ao número médio de saídas, em cada estado.

Admitindo taxas de chegada λ e de serviço μ , constantes e independentes do estado do sistema, temos o número médio de clientes no sistema e o número médio de clientes na fila dados, respectivamente, por

$$L_s = \sum_{n=0}^{\infty} n P_n$$

e

$$L_q = \sum_{n=S+1}^{\infty} (n - S) P_n,$$

onde S representa o número de postos de serviço em paralelo.

Quando um processo está no estado estacionário, são aplicáveis as fórmulas de Little, apresentadas por este nos anos 60 (cf. [41]), que afirmam que:

$$L_s = \lambda W_s \tag{6}$$

e

$$L_q = \lambda W_q. \tag{7}$$

Estas equações também são válidas, com uma pequena adaptação, se λ depender do estado do sistema. Neste caso deve-se substituir λ por λ' , onde $\lambda' = \sum_{n=0}^{\infty} \lambda_n P_n$, (cf [19]).

Note-se que, quando a taxa de atendimento é μ , igual para todos os clientes, então

$$W_s = W_q + \frac{1}{\mu}, \tag{8}$$

qualquer que seja o número de servidores.

De (6) e de (8) resulta que,

$$L_s = L_q + \frac{\lambda}{\mu}. \tag{9}$$

A modelação de filas de espera através de processos de vida e morte supõe a existência de estados estacionários. Excluí-se assim as filas que tendem para o infinito porque a taxa de chegada excede a capacidade de atendimento, e as filas que tenham a taxa de chegada variável ao longo do tempo e inícios de funcionamento com números anormais de clientes.

3.3.2 Algumas notas

As chegadas ocorrem no tempo e no espaço de acordo com as leis de probabilidade. Por isso é necessário conhecer a distribuição de probabilidade que descreve as chegadas de clientes. A mais conveniente é a distribuição de Poisson, tomando assim os tempos entre as chegadas exponencialmente distribuídos, como vimos no final da secção 2.4.

Note-se que a taxa de chegadas pode variar ao longo do tempo.

Não é difícil mostrar que se os clientes chegarem de acordo com um processo de Poisson com parâmetro λ , e se puderem ser divididos em diferentes tipos, então a chegada de cada um dos clientes de cada um dos tipos, onde p_i é a probabilidade desse cliente pertencer ao tipo i , segue ainda uma distribuição de Poisson com parâmetro $\lambda_i = p_i \lambda$.

No processo de chegada também é necessário conhecer a distribuição de probabilidade do atendimento. A mais conveniente é a distribuição exponencial que, além de ser um bom modelo para a realidade, é matematicamente “tratável”.

Consideremos agora T_1, \dots, T_n variáveis aleatórias independentes com distribuição exponencial e parâmetros $\alpha_1, \dots, \alpha_n$. A variável $U = \min \{T_1, \dots, T_n\}$, que representa o tempo até o primeiro acontecimento que ocorra de entre os n , segue uma distribuição exponencial com parâmetro $\alpha = \sum_{i=1}^n \alpha_i$. De facto, para $t > 0$

$$\begin{aligned} P[U > t] &= P[\min \{T_1, \dots, T_n\} > t] \\ &= P[T_1 > t, \dots, T_n > t] \\ &= P[T_1 > t] \dots P[T_n > t] \\ &= e^{-\alpha_1 t} \dots e^{-\alpha_n t} \\ &= e^{-(\alpha_1 + \dots + \alpha_n)t}. \end{aligned}$$

Esta propriedade permite modelar sistemas com k servidores idênticos com parâmetro μ , operando em paralelo, como um único servidor com parâmetro $k\mu$ (supondo que o primeiro cliente da fila é atendido no momento em que um servidor é libertado).

4 Modelos $M/M/a/b/c$

O modelo mais profundamente estudado e o mais simples como já foi referido, é o que tem chegadas dos clientes ocorrendo segundo um processo de Poisson e o tempo de serviço com distribuição exponencial. Este modelo denota-se por $M/M/k$, onde $k \in \{\mathbb{N} \cup \infty\}$.

Assumamos que a taxa das chegadas dos clientes é λ e que o tempo de serviço tem distribuição exponencial com parâmetro μ . Uma forma de definir o modelo referido (cf. [42]) é a seguinte (para $h \downarrow 0$),

$$P[\text{"Haver uma chegada em } [t, t+h)\text{"}] = \lambda h + o(h),$$

$$P[\text{"Não haver uma chegada em } [t, t+h)\text{"}] = 1 - \lambda h + o(h),$$

$$P\left[\begin{array}{l} \text{"O serviço ser completado em } [t, t+h)\text{"} \\ \text{"O serviço está a ser realizado no instante } t\text{"} \end{array}\right] = \mu h + o(h),$$

$$P\left[\begin{array}{l} \text{"O serviço não ser completado em } [t, t+h)\text{"} \\ \text{"O serviço está a ser realizado no instante } t\text{"} \end{array}\right] = 1 - \mu h + o(h),$$

Ao longo deste capítulo designamos por $X(t)$ o número de clientes no sistema no instante t ($t > 0$).

4.1 Modelo $M/M/1$

Este modelo corresponde ao modelo básico onde o sistema tem uma distribuição das chegadas de Poisson e dos tempos de atendimento exponencial e contém um só servidor, a capacidade do sistema e da população são infinitas e a disciplina é a mais comum, correspondendo a quem entra primeiro no sistema ser o primeiro a ser atendido e a sair. Este pode ser representado também como $M/M/1/\infty/\infty/FIFO$ ou $M/M/1/\infty/\infty/FCFS$. Um exemplo deste modelo pode ser de uma loja com um único balcão de atendimento.

A probabilidade de chegar alguém que tem de aguardar para ser servido é

$$\begin{aligned} P[X(t+h) = k+1 | X(t) = k] &= [\lambda h + o(h)] \times [1 - \mu h + o(h)] \\ &= \lambda h + o(h) \text{ com } k = 0, 1, \dots \end{aligned}$$

A probabilidade de não chegar ninguém e haver um serviço que foi completado é

$$\begin{aligned} P[X(t+h) = k-1 | X(t) = k] &= [1 - \lambda h + o(h)] \times [\mu h + o(h)] \\ &= \mu h + o(h). \end{aligned}$$

Ou seja, $\{X(t)\}$ é um processo de vida e de morte com parâmetros λ e μ , respectivamente.

Sabendo que o número médio de clientes no sistema (L_s) é, por definição, a média de todos os estados possíveis do sistema ponderados pelas respectivas probabilidades de ocorrência, temos que

$$L_s = \sum_{n=0}^{\infty} n P_n.$$

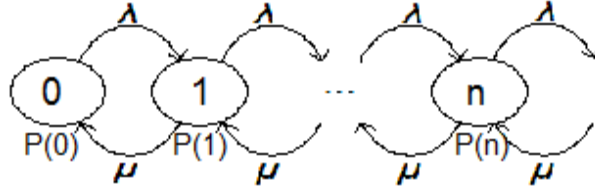


Figura 1: Modelo M/M/1

Note-se que este sistema poderá ser representado através da Figura 1, onde os números representam os estados possíveis, e as setas as transições possíveis.

No estado 0, o *número de entradas* por unidade de tempo é μP_1 , que ocorre na passagem do estado 1 para o estado 0 por saída do sistema de um cliente atendido, e o *número de saídas* por unidade de tempo é λP_0 , que ocorre na passagem do estado 0 para o estado 1 por chegada de um cliente ao sistema.

Para que o estado de equilíbrio ou estacionário ocorra é necessário que o número de entradas por unidade de tempo e o número de saídas por unidade de tempo sejam iguais, isto é, que,

$$\mu P_1 = \lambda P_0 \iff P_1 = \frac{\lambda}{\mu} P_0. \quad (10)$$

No estado 1 o *número de entradas* por unidade de tempo é $\lambda P_0 + \mu P_2$, que acontece na passagem do estado 0 com a chegada de um cliente ao sistema ou na passagem do estado 2 por saída do sistema de um cliente atendido, e o *número de saídas* por unidade de tempo é $\lambda P_1 + \mu P_1$, que acontece na passagem para o estado 2 por chegada de um cliente ou para o estado 0 por saída do sistema de um cliente atendido.

Igualando o número de entradas com o número de saídas, temos a equação de equilíbrio,

$$\lambda P_0 + \mu P_2 = \lambda P_1 + \mu P_1 \iff P_2 = \frac{\lambda P_1 + \mu P_1 - \lambda P_0}{\mu}.$$

Da equação (10) sabemos que $P_0 = \frac{\mu}{\lambda} P_1$. Usando esse facto temos que:

$$\begin{aligned} P_2 &= \frac{\lambda P_1 + \mu P_1 - \lambda P_0}{\mu} = \frac{\lambda P_1 + \mu P_1 - \lambda \frac{\mu}{\lambda} P_1}{\mu} = \frac{\lambda P_1 + \mu P_1 - \mu P_1}{\mu} \\ &= \frac{\lambda}{\mu} P_1. \end{aligned}$$

Usando novamente a equação (10) obtemos

$$P_2 = \frac{\lambda}{\mu} P_1 = \frac{\lambda}{\mu} \left(\frac{\lambda}{\mu} P_0 \right) \iff P_2 = \left(\frac{\lambda}{\mu} \right)^2 P_0. \quad (11)$$

No estado 2 com um raciocínio análogo verificamos que o *número de entradas* por unidade de tempo é $\lambda P_1 + \mu P_3$ e o *número de saídas* por unidade de tempo é $\lambda P_2 + \mu P_2$.

Então o estado de equilíbrio corresponde a

$$\lambda P_1 + \mu P_3 = \lambda P_2 + \mu P_2 \iff P_3 = \frac{\lambda P_2 + \mu P_2 - \lambda P_1}{\mu}.$$

Da equação (11) sabemos que $P_1 = \frac{\mu}{\lambda} P_2$, de onde resulta que

$$\begin{aligned} P_3 &= \frac{\lambda P_2 + \mu P_2 - \lambda P_1}{\mu} = \frac{\lambda P_2 + \mu P_2 - \lambda \frac{\mu}{\lambda} P_2}{\mu} = \frac{\lambda P_2 + \mu P_2 - \mu P_2}{\mu} \\ &= \frac{\lambda}{\mu} P_2. \end{aligned}$$

E tendo em conta novamente a equação (11) temos

$$P_3 = \frac{\lambda}{\mu} P_2 = \frac{\lambda}{\mu} \left(\frac{\lambda}{\mu} \right)^2 P_0 \iff P_3 = \left(\frac{\lambda}{\mu} \right)^3 P_0.$$

Estudando os estados anteriores chegamos facilmente à conclusão que

$$P_n = \left(\frac{\lambda}{\mu} \right)^n P_0.$$

Se designarmos por ρ o quociente $\frac{\lambda}{\mu}$, conhecido por "taxa de ocupação", vem que

$$P_n = \rho^n P_0. \quad (12)$$

Impondo a condição

$$\sum_{n=0}^{\infty} P_n = 1$$

resulta que

$$\sum_{n=0}^{\infty} P_n = 1 \iff \sum_{n=0}^{\infty} \rho^n P_0 = 1 \iff \frac{1}{P_0} = \sum_{n=0}^{\infty} \rho^n.$$

O somatório anterior corresponde a uma progressão geométrica de razão ρ , que será convergente se $\rho < 1$

$$\frac{1}{P_0} = \sum_{n=0}^{\infty} \rho^n \iff \frac{1}{P_0} = \frac{1}{1-\rho} \iff P_0 = 1-\rho$$

Substituindo P_0 da equação anterior na equação (12) temos que

$$P_n = \rho^n (1-\rho) = \left(\frac{\lambda}{\mu} \right)^n \left(1 - \frac{\lambda}{\mu} \right), \quad (\rho < 1, \lambda < \mu), \quad (13)$$

isto é, P_n tem distribuição de uma v.a. geométrica de parâmetro $\frac{\lambda}{\mu}$.

O processo acima descrito poderia ter sido observado. De facto, considerando que este processo pode ser visto como um processo de nascimento e de morte com parâmetros de nascimento dados por $\lambda_n = \lambda, \forall n \in \mathbb{N}_0$ e parâmetros de morte dados por $\mu_n = \mu, \forall n \in \mathbb{N}_1$, sabemos (cf. Subsecção 2.4.1) que a distribuição estacionária para este processo é dada por

$$P_n = \lim_{t \rightarrow +\infty} P[X(t) = n] = \frac{\theta_n}{\sum_{n \geq 0} \theta_n},$$

onde

$$\theta_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n}.$$

desde que $\sum_{n \geq 0} \theta_n < +\infty$.

No modelo aqui considerado,

$$\theta_n = \left(\frac{\lambda}{\mu}\right)^n$$

donde, se $\frac{\lambda}{\mu} < 1$,

$$\sum_{n \geq 0} \theta_n = \frac{1}{1 - \frac{\lambda}{\mu}},$$

e

$$\begin{aligned} P_n &= \frac{\theta_n}{\sum_{n \geq 0} \theta_n} = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) \\ &= \rho^n (1 - \rho). \end{aligned}$$

Na Figura 2 representamos P_n , $n = 0, 1, 5$ e 20 , como funções da taxa de ocupação, ρ , e representamos P_n , $n = 10, 20, 30$ e 40 , numa escala diferente na Figura 3.

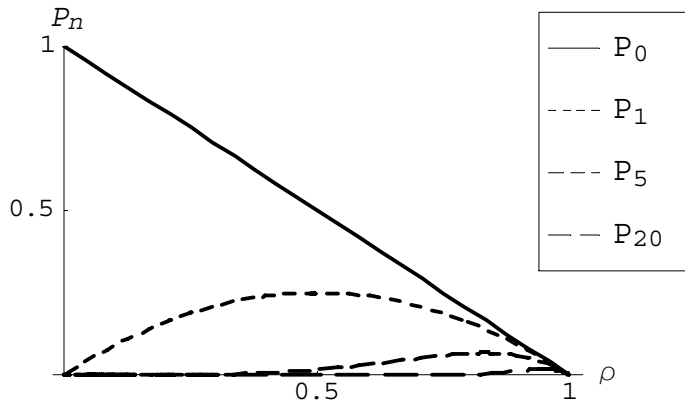


Figura 2: Probabilidade de existirem n clientes no sistema (P_n) numa escala de 0 a 1 em função da taxa de ocupação (ρ)

Através da Figura 2 e da Figura 3 podemos observar que quando a taxa de ocupação tende para 1, as probabilidades tendem para 0, e que quanto maior for o número de clientes, mas a sua probabilidade se aproxima de 0, independentemente da taxa de ocupação.

Usando estes resultados podemos chegar ao conjunto de características que nos interessam para o modelo $M/M/1/\infty/\infty/FIFO$.

O número médio de clientes no sistema (L_s), quando este se encontra em situação de equilíbrio, é

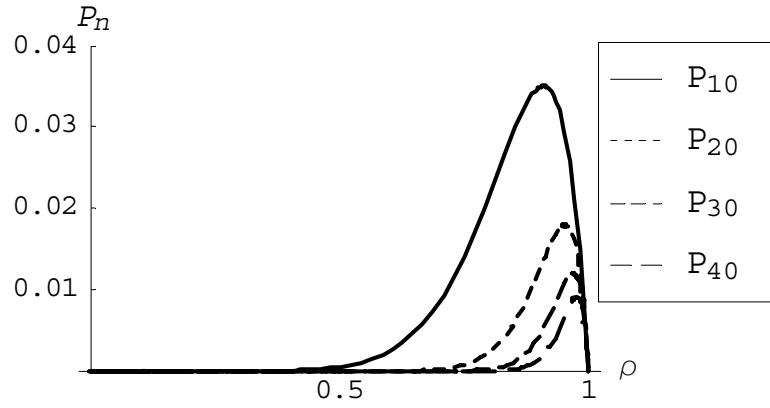


Figura 3: Probabilidade de existirem n clientes no sistema (P_n) numa escala de 0 a 0.04 em função da taxa de ocupação (ρ)

$$L_s = \sum_{n=0}^{\infty} n P_n.$$

Substituindo a expressão obtida em (13) para P_n na equação anterior temos

$$L_s = \sum_{n=0}^{\infty} n P_n = \sum_{n=0}^{\infty} n \rho^n (1 - \rho).$$

Desenvolvendo o somatório anterior usando as regras dos somatórios e das derivadas, temos que, para $\rho < 1$,

$$\begin{aligned} L_s &= \sum_{n=0}^{\infty} n \rho^n (1 - \rho) = (1 - \rho) \rho \sum_{n=1}^{\infty} n \rho^{n-1} = (1 - \rho) \rho \frac{d}{d\rho} \sum_{n=0}^{\infty} \rho^n \\ &= (1 - \rho) \rho \frac{d}{d\rho} \left(\frac{1}{1 - \rho} \right) = (1 - \rho) \rho \frac{1}{(1 - \rho)^2} \\ &= \frac{\rho}{1 - \rho}. \end{aligned}$$

Usando o facto de $\rho = \frac{\lambda}{\mu}$,

$$\begin{aligned} L_s &= \frac{\rho}{1 - \rho} = \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}} \\ &= \frac{\lambda}{\mu - \lambda}. \end{aligned} \tag{14}$$

Usando (14) podemos calcular o comprimento médio da fila (L_q) e o tempo médio de espera no sistema (W_s).

Em primeiro lugar vamos determinar L_q . Com ajuda da igualdade (9), temos que

$$L_s = L_q + \frac{\lambda}{\mu} \iff L_q = L_s - \frac{\lambda}{\mu}.$$

Substituindo L_s da equação (14) na equação anterior, obtemos

$$\begin{aligned} L_q &= \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu} = \frac{\lambda\mu - \lambda(\mu - \lambda)}{\mu(\mu - \lambda)} \\ &= \frac{\lambda^2}{\mu(\mu - \lambda)}. \end{aligned} \quad (15)$$

Alternativamente, poderíamos escrever L_q em função de ρ

$$\begin{aligned} L_q &= \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\left(\frac{\lambda}{\mu}\right)^2}{1 - \frac{\lambda}{\mu}} \\ &= \frac{\rho^2}{1 - \rho}. \end{aligned}$$

Na Figura 4 representamos L_s , o tempo de espera de um cliente no sistema, conjuntamente com L_q , o tempo de espera de um cliente na fila, em função de ρ , a taxa de ocupação.

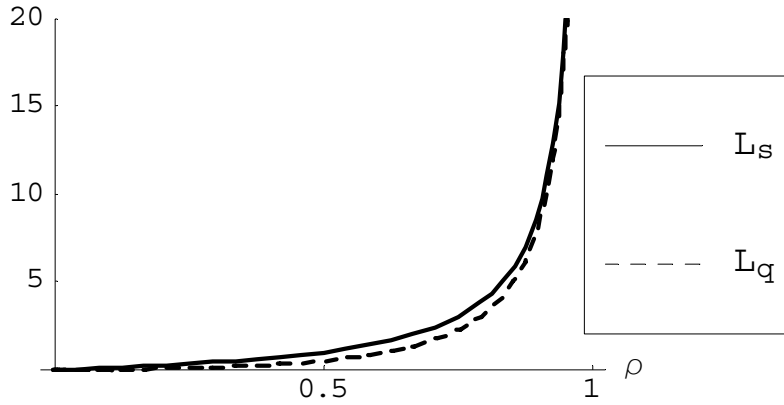


Figura 4: Número médio de clientes no sistema (L_s) e na fila (L_q) em função da taxa de ocupação (ρ)

Como já era de esperar, a Figura 4 mostra que consoante aumenta a taxa de ocupação, mais clientes ficam no sistema à espera, de saírem do sistema, e quando a taxa de ocupação, ρ converge para 1, o número de clientes no sistema tende para o infinito, tanto em L_s como em L_q .

Vamos agora determinar W_s usando as igualdades (6) e (14):

$$\begin{aligned} L_s &= \lambda W_s \\ \iff W_s &= \frac{L_s}{\lambda} = \frac{\frac{\lambda}{\mu - \lambda}}{\lambda} = \frac{1}{\mu - \lambda}, \end{aligned}$$

ou, alternativamente,

$$W_s = \frac{1}{\mu(1-\rho)}.$$

Na Figura 5 representamos W_s como função da taxa de ocupação para os casos $\mu = 0.5$, $\mu = 1$, $\mu = 5$ e $\mu = 50$.

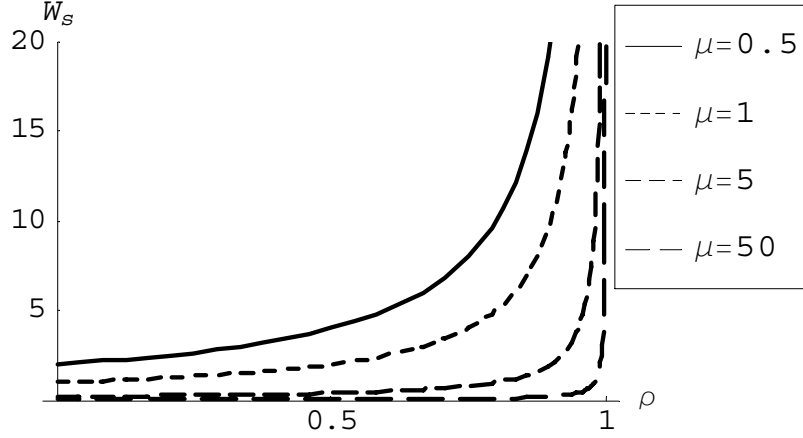


Figura 5: Tempo médio de espera no sistema (W_s) em função da taxa de ocupação (ρ)

Podemos concluir da Figura 5 que, como esperado, independentemente da taxa de serviço, quando a taxa de ocupação aumenta, o tempo médio de espera dos clientes no sistema aumenta e tende para o infinito, quando ρ converge para 1. Em relação à taxa de serviço, podemos concluir que quanto maior for a taxa de serviço menos tempo os clientes ficam no sistema à espera de saírem do sistema.

Falta-nos calcular o tempo médio de espera na fila (W_q) que será facilmente descoberto através de (7) e de (15).

$$\begin{aligned} L_q &= \lambda W_q \\ W_q &= \frac{L_q}{\lambda} = \frac{\frac{\lambda^2}{\mu(\mu-\lambda)}}{\lambda} = \frac{\lambda}{\mu(\mu-\lambda)} \\ &= \frac{\rho}{\mu(1-\rho)}. \end{aligned}$$

Como no caso anterior representamos W_q em função de ρ para os casos $\mu = 1$, $\mu = 5$ e $\mu = 50$ (ver Figura 6).

Como seria de esperar, os gráficos de W_q e W_s , Figura 6 e Figura 5, respectivamente, são semelhantes.

Os valores médios anteriores foram deduzidos através da expressão da distribuição estacionária $\{P_n\}_{n \in \mathbb{N}_0}$ e da fórmula de Little. No entanto neste modelo, dada a sua simplicidade, é possível facilmente ir mais longe e determinar as distribuições das variáveis que registam o tempo que um cliente está no sistema (fila), T_s (T_q).

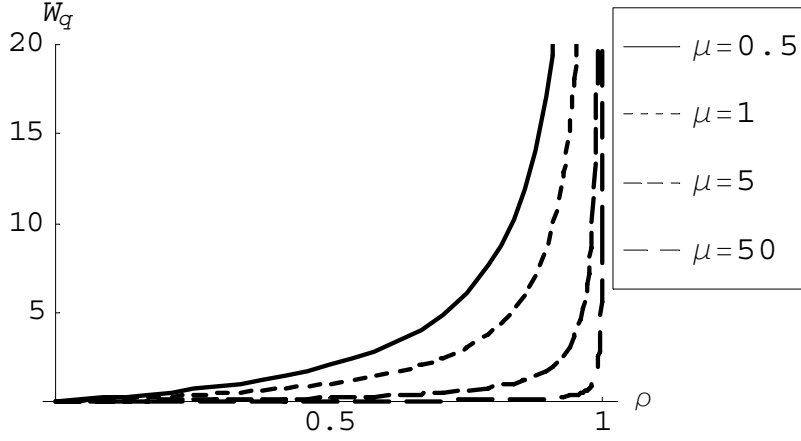


Figura 6: Tempo médio de espera na fila (W_q) em função da taxa de ocupação (ρ)

De facto, se um cliente chega ao sistema quando neste existem n indivíduos, o seu tempo de espera total (incluindo o tempo do seu serviço), isto é, tempo de espera que corresponde a $n + 1$ clientes, W_s , tem distribuição gama de parâmetros $(\lambda, n + 1)$, ou seja,

$$\begin{aligned} F_{T_s} | \text{"n clientes à frente"} (t) &= P[T_s \leq t | \text{"n clientes à frente"}] \\ &= \int_0^t \frac{(\mu x)^n}{n!} \mu e^{-\mu x} dx \quad (t \geq 0). \end{aligned} \quad (16)$$

Usando o teorema da probabilidade total, cf. (1), obtemos, para $t \geq 0$,

$$\begin{aligned} F_{T_s}(t) &= P[T_s \leq t] = \sum_{n=0}^{+\infty} P[T_s \leq t | \text{"n clientes à frente"}] P_n \\ &= \sum_{n=0}^{+\infty} \left[\int_0^t \frac{(\mu x)^n}{n!} \mu e^{-\mu x} dx \left(\frac{\lambda}{\mu} \right)^n \left(1 - \frac{\lambda}{\mu} \right) \right] \\ &= \left(1 - \frac{\lambda}{\mu} \right) \mu \sum_{n=0}^{+\infty} \left[\int_0^t \frac{(\mu x)^n}{n!} \left(\frac{\lambda}{\mu} \right)^n e^{-\mu x} dx \right] \\ &= (\mu - \lambda) \sum_{n=0}^{+\infty} \left[\int_0^t \frac{(\lambda x)^n}{n!} e^{-\mu x} dx \right]. \end{aligned}$$

Tendo em conta que a função integranda é não negativa, resulta que

$$\begin{aligned} F_{T_s}(t) &= (\mu - \lambda) \int_0^t \left(\sum_{n=0}^{+\infty} \frac{(\lambda x)^n}{n!} \right) e^{-\mu x} dx = (\mu - \lambda) \int_0^t e^{\lambda x} e^{-\mu x} dx = (\mu - \lambda) \int_0^t e^{\lambda x - \mu x} dx \\ &= (\mu - \lambda) \left[\frac{e^{\lambda x - \mu x}}{\lambda - \mu} \right]_0^t = (\mu - \lambda) \left[\frac{e^{(\lambda - \mu)t} - e^0}{\lambda - \mu} \right] = (\mu - \lambda) \left[\frac{1 - e^{(\lambda - \mu)t}}{\mu - \lambda} \right] \\ &= 1 - e^{(\lambda - \mu)t}, \end{aligned}$$

isto é, $T_s \sim \text{Exponencial}(\mu - \lambda)$. A probabilidade do tempo de espera na fila exceder t , é

$$\begin{aligned} P(T_q > t) &= \frac{\lambda}{\mu} P(T_s > t) \\ &= \frac{\lambda}{\mu} e^{-(\mu-\lambda)t} \end{aligned}$$

pois de (7) e de (6)

$$\begin{aligned} W_q &= \frac{L_q}{L_s} W_s = \frac{\frac{\lambda^2}{\mu(\mu-\lambda)}}{\frac{\lambda}{\mu-\lambda}} W_s \\ &= \frac{\lambda}{\mu} W_s. \end{aligned}$$

Vamos então agora calcular a probabilidade de existirem no sistema (em situação de equilíbrio) k ou mais clientes. Seja N_0 número de clientes no sistema, supondo que este se encontra numa situação de equilíbrio. Temos

$$P(N \geq k) = \sum_{n=k}^{\infty} P_n$$

Substituindo a expressão de P_n obtida em (13) na equação anterior temos:

$$\begin{aligned} P(N \geq k) &= \sum_{n=k}^{\infty} \rho^n (1 - \rho) = (1 - \rho) \sum_{n=k}^{\infty} \rho^n \\ &= (1 - \rho) \rho^k \sum_{n=0}^{\infty} \rho^n = (1 - \rho) \rho^k \frac{1}{1 - \rho} \\ &= \rho^k. \end{aligned}$$

A probabilidade do tempo de espera na fila ser zero, $P(T_q = 0)$, é dada por,

$$P(T_q = 0) = P_0 = 1 - \rho,$$

representada graficamente na Figura 7.

Como era de esperar, podemos verificar da Figura 7, a probabilidade de um cliente chegar a um sistema e não ter de esperar, diminui com o aumento da taxa de ocupação.

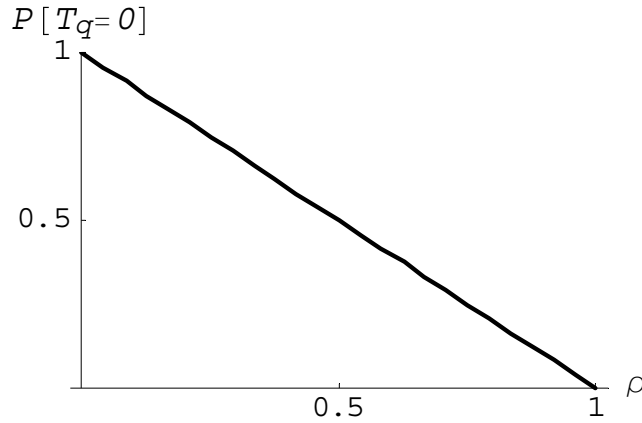


Figura 7: Probabilidade de o tempo de espera na fila ser zero $P[T_q = 0]$ em função de ρ

Resumindo as características do modelo $M/M/1$, temos o seguinte quadro:

Probabilidade de ocorrência do estado 0	$P_0 = 1 - \rho$
Probabilidade de ocorrência do estado n	$P_n = \rho^n P_0$
Número médio de clientes no sistema	$L_s = \frac{\rho}{1-\rho}$
Comprimento médio da fila	$L_q = \frac{\rho^2}{1-\rho}$
Tempo médio de espera no sistema	$W_s = \frac{1}{\mu(1-\rho)}$
Tempo médio de espera na fila	$W_q = \frac{\rho}{\mu(1-\rho)}$
Probabilidade de existirem no sistema k ou mais clientes	$P(N \geq k) = \rho^k$
Probabilidade de o tempo de espera na fila ser zero	$P(T_q = 0) = P_0$
Probabilidade de o tempo de espera na fila exceder t	$P(T_q > t) = \frac{\lambda}{\mu} e^{(\lambda-\mu)t}$
Probabilidade de o tempo gasto no sistema exceder t	$P(T_s > t) = e^{(\lambda-\mu)t}$
Taxa de ocupação	$\rho = \frac{\lambda}{\mu}$

Quadro 1: Sumário das características do modelo $M/M/1$

4.2 Modelo $M/M/S$

Este modelo difere do anterior apenas no número de servidores disponíveis: S . Temos assim um modelo em que o número de servidores é S , o sistema tem uma distribuição das chegadas de Poisson e dos tempos de atendimento exponencial, a capacidade do sistema e da população são infinitas e a disciplina corresponde a quem entra primeiro no sistema ser o primeiro a ser atendido e a sair. Temos como exemplo um supermercado com S balcões de atendimento.

Este modelo, em que existem S servidores em paralelo, corresponde a um processo de vida e morte. A *taxa de chegada* λ é independente do número de clientes e a *taxa de serviço* μ é igual para todos os serviços, sendo a taxa de atendimento dada por $S\mu$. E por conseguinte a *taxa de ocupação* ρ é $\rho = \frac{\lambda}{S\mu}$.

A taxa de entrada em cada estado é sempre igual a λ e a taxa de saída varia com o estado, pois se os servidores não estiverem todos ocupados, a distribuição exponencial do intervalo de tempo entre saídas de clientes atendidos tem parâmetro $k\mu$ (com $k < S$), sendo k o número de servidores ocupados nesse estado, e se todos os servidores estiverem ocupados, a distribuição exponencial do intervalo de tempo entre saídas de clientes atendidos tem parâmetro $S\mu$, ou seja,

$$\mu_n = \begin{cases} n\mu & \text{se } n = 1, \dots, S \\ S\mu & \text{se } n > S \end{cases}.$$

Este sistema pode ser representado através da Figura 8.

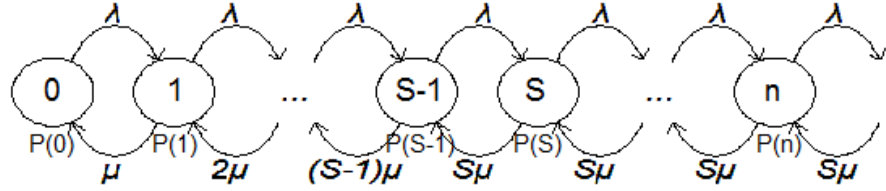


Figura 8: Modelo $M/M/S$

Assim sendo,

$$\begin{aligned} \theta_n &= \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} = \begin{cases} \frac{\lambda^n}{n! \mu^n} & \text{se } n = 0, 1, \dots, S \\ \frac{\lambda^n}{S! S^{n-S} \mu^n} & \text{se } n > S \end{cases} \\ &= \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n & \text{se } n = 0, 1, \dots, S \\ \frac{1}{S! S^{n-S}} \left(\frac{\lambda}{\mu}\right)^n & \text{se } n > S \end{cases}. \end{aligned}$$

Quando $n > 1$ e $\sum_{n \geq 0} \theta_n < \infty$, temos então que

$$\begin{aligned} P_n &= \lim_{t \rightarrow +\infty} P[X(t) = n] = \frac{\theta_n}{\sum_{n \geq 0} \theta_n} = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n P_0 & \text{se } n = 0, 1, \dots, S \\ \frac{S^n}{S! S^{n-S}} \left(\frac{\lambda}{\mu}\right)^n P_0 & \text{se } n > S \end{cases} \\ &= \begin{cases} \frac{S^n}{n!} \rho^n P_0 & \text{se } n = 0, 1, \dots, S \\ \frac{S^n}{S! S^{n-S}} \rho^n P_0 & \text{se } n > S \end{cases} \\ &= \begin{cases} \frac{S^n}{n!} \rho^n P_0 & \text{se } n = 0, 1, \dots, S \\ \frac{S^n}{S!} \rho^n P_0 & \text{se } n > S \end{cases}. \end{aligned}$$

onde

$$P_0 = \frac{1}{\sum_{n \geq 0} \theta_n}.$$

Ora, se $S\mu > \lambda$, isto é, se $\rho < 1$, temos que

$$\begin{aligned}
 \frac{1}{P_0} &= \sum_{n \geq 0} \theta_n = \sum_{0 \leq n \leq S} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \sum_{n \geq S+1} \frac{1}{S!} \left(\frac{\lambda}{S\mu}\right)^{n-S} \left(\frac{\lambda}{\mu}\right)^S \\
 &= \sum_{0 \leq n \leq S} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{S!} \left(\frac{\lambda}{\mu}\right)^S \sum_{n \geq 0} \left(\frac{\lambda}{S\mu}\right)^{n-S+S+1} \\
 &= \sum_{0 \leq n \leq S} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{S!} \left(\frac{\lambda}{\mu}\right)^S \left(\frac{\lambda}{S\mu}\right) \left(\frac{1}{1 - \frac{\lambda}{S\mu}}\right) \\
 &= \sum_{0 \leq n \leq S-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{S!} \left(\frac{\lambda}{\mu}\right)^S \left(\frac{1}{1 - \frac{\lambda}{S\mu}}\right). \tag{17}
 \end{aligned}$$

Podemos concluir que

$$\begin{aligned}
 P_0 &= \frac{1}{\sum_{0 \leq n \leq S-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{S!} \left(\frac{\lambda}{\mu}\right)^S \left(\frac{1}{1 - \frac{\lambda}{S\mu}}\right)} \\
 &= \frac{1}{\sum_{0 \leq n \leq S-1} \frac{S^n}{n!} \rho^n + \frac{S^S}{S!} \rho^S \left(\frac{1}{1-\rho}\right)}.
 \end{aligned}$$

Supondo que $S = 10$, representamos na Figura 9 P_n , para $n = 0, 1, 5$ e 10 .

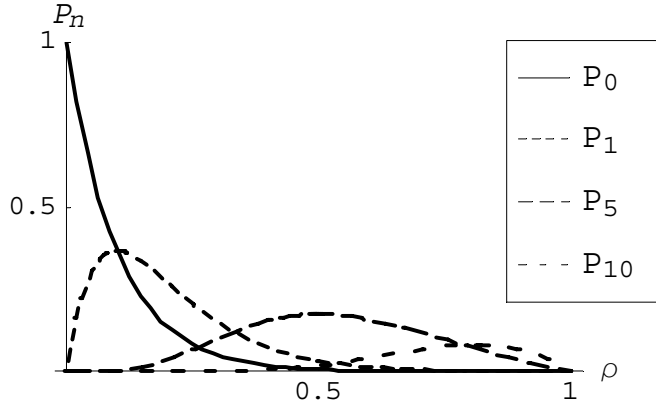


Figura 9: Probabilidade de existirem n clientes no sistema (P_n) em função da taxa de ocupação (ρ), com $S = 10$

Da Figura 9, podemos observar que, exceptuando P_0 , a probabilidade começa em 0, aumenta e volta a convergir para 0 quando a taxa de ocupação converge para 1. Relativamente a P_0 , como seria de esperar, neste caso extremo em que a taxa de ocupação é 0 (que na realidade não se concretiza pois $\rho > 0$), a probabilidade do sistema estar no estado 0 é 1, mas tende para 0, quando ρ converge para 1.

Das igualdades anteriores, podemos tirar algumas conclusões para alguns somatórios que nos serão úteis mais tarde.

De (17) temos

$$\begin{aligned} \sum_{0 \leq n \leq S} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n &= \frac{1}{P_0} - \frac{1}{S!} \left(\frac{\lambda}{\mu} \right)^S \left(\frac{\lambda}{S\mu} \right) \left(\frac{1}{1 - \frac{\lambda}{S\mu}} \right) \\ &= \frac{1}{P_0} - \frac{P_S}{P_0} \left(\frac{\lambda}{S\mu} \right) \left(\frac{1}{1 - \frac{\lambda}{S\mu}} \right). \end{aligned}$$

Consequentemente,

$$\begin{aligned} \sum_{0 \leq n \leq S-1} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n &= \frac{1}{P_0} - \frac{P_S}{P_0} \left(\frac{\lambda}{S\mu} \right) \left(\frac{1}{1 - \frac{\lambda}{S\mu}} \right) - \frac{1}{S!} \left(\frac{\lambda}{\mu} \right)^S \\ &= \frac{1}{P_0} - \frac{P_S}{P_0} \left(\frac{\lambda}{S\mu - \lambda} \right) - \frac{P_S}{P_0} = \frac{1}{P_0} - \frac{P_S}{P_0} \left(\frac{\lambda}{S\mu - \lambda} - 1 \right) \\ &= \frac{1}{P_0} - \frac{P_S}{P_0} \left(\frac{\lambda - S\mu + \lambda}{S\mu - \lambda} \right) = \frac{1}{P_0} - \frac{P_S}{P_0} \left(\frac{S\mu}{S\mu - \lambda} \right) \\ &= \frac{1}{P_0} - \frac{P_S}{P_0} \left(\frac{1}{1 - \frac{\lambda}{S\mu}} \right) = \frac{1}{P_0} \left(1 - \frac{P_S}{1 - \frac{\lambda}{S\mu}} \right). \end{aligned}$$

E ainda temos que

$$\begin{aligned} \sum_{0 \leq n \leq S} P_n &= \sum_{0 \leq n \leq S} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0 = P_0 \left[\frac{1}{P_0} - \frac{P_S}{P_0} \left(\frac{\lambda}{S\mu} \right) \left(\frac{1}{1 - \frac{\lambda}{S\mu}} \right) \right] \\ &= 1 - \left(\frac{\lambda}{S\mu} \right) \left(\frac{P_S}{1 - \frac{\lambda}{S\mu}} \right). \end{aligned}$$

Suponhamos que $S\mu > \lambda$, isto é, que $\rho = \frac{\lambda}{S\mu} < 1$. De seguida vamos calcular o valor esperado de número de clientes no sistema. Sabemos que

$$L_s = \sum_{n=0}^{\infty} nP_n = \sum_{0 \leq n \leq S} nP_n + \sum_{n \geq S+1} nP_n.$$

Determinando em primeiro lugar o primeiro somatório obtemos,

$$\begin{aligned} \sum_{0 \leq n \leq S} nP_n &= \sum_{1 \leq n \leq S} n \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0 = \frac{\lambda}{\mu} P_0 \sum_{0 \leq n \leq S-1} \frac{\left(\frac{\lambda}{\mu} \right)^n}{n!} \\ &= \frac{\lambda}{\mu} \left(1 - \frac{P_S}{1 - \frac{\lambda}{S\mu}} \right). \end{aligned}$$

Relativamente ao segundo somatório temos

$$\begin{aligned}
\sum_{n \geq S+1} n P_n &= \sum_{n \geq S+1} n \frac{1}{S!} \left(\frac{\lambda}{S\mu} \right)^{n-S} \left(\frac{\lambda}{\mu} \right)^S P_0 = \frac{1}{S!} \left(\frac{\lambda}{\mu} \right)^S P_0 \sum_{n \geq 1} (n+S) \left(\frac{\lambda}{S\mu} \right)^n \\
&= \frac{1}{S!} \left(\frac{\lambda}{\mu} \right)^S P_0 \left[\sum_{n \geq 1} n \left(\frac{\lambda}{S\mu} \right)^n + S \sum_{n \geq 1} \left(\frac{\lambda}{S\mu} \right)^n \right] \\
&= \frac{1}{S!} \left(\frac{\lambda}{\mu} \right)^S P_0 \left[\left(\frac{\lambda}{S\mu} \right) \left[\frac{d}{dx} \sum_{n \geq 0} x^n \right]_{x=\frac{\lambda}{S\mu}} + S \frac{\frac{\lambda}{S\mu}}{1 - \frac{\lambda}{S\mu}} \right] \\
&= P_S \left[\frac{\lambda}{S\mu} \frac{1}{\left(1 - \frac{\lambda}{S\mu}\right)^2} + \frac{\lambda}{\mu} \frac{1}{1 - \frac{\lambda}{S\mu}} \right].
\end{aligned}$$

Juntando os dois somatórios temos que

$$\begin{aligned}
L_s &= \frac{\lambda}{\mu} \left(1 - \frac{P_S}{1 - \frac{\lambda}{S\mu}} \right) + P_S \left[\frac{\lambda}{S\mu} \frac{1}{\left(1 - \frac{\lambda}{S\mu}\right)^2} + \frac{\lambda}{\mu} \frac{1}{1 - \frac{\lambda}{S\mu}} \right] \\
&= \frac{\lambda}{\mu} - \left(\frac{\lambda}{\mu} \right) \frac{P_S}{1 - \frac{\lambda}{S\mu}} + \frac{\lambda}{S\mu} \frac{P_S}{\left(1 - \frac{\lambda}{S\mu}\right)^2} + \frac{\lambda}{\mu} \frac{P_S}{1 - \frac{\lambda}{S\mu}} \\
&= \frac{\lambda}{\mu} + \frac{\lambda}{S\mu} \frac{P_S}{\left(1 - \frac{\lambda}{S\mu}\right)^2}.
\end{aligned}$$

Escrevendo $\rho = \frac{\lambda}{S\mu}$ vem

$$\begin{aligned}
L_s &= \frac{\lambda}{\mu} + \rho \frac{P_S}{(1 - \rho)^2} \\
&= S\rho + \rho \frac{P_S}{(1 - \rho)^2}.
\end{aligned}$$

De (9) resulta

$$L_s = L_q + \frac{\lambda}{\mu},$$

donde podemos concluir que

$$L_q = \frac{\lambda}{S\mu} \frac{P_S}{\left(1 - \frac{\lambda}{S\mu}\right)^2}, \quad (18)$$

ou, equivalentemente, que

$$L_q = \rho \frac{P_S}{(1 - \rho)^2}.$$

Supondo que $S = 10$, representamos graficamente L_q em simultâneo com L_s na Figura 10.

Fazendo uma análise à Figura 10, podemos observar que o comprimento da fila mantém-se muito próximo do valor zero para valores de ρ inferiores a $\frac{1}{2}$, e cresce à medida que a

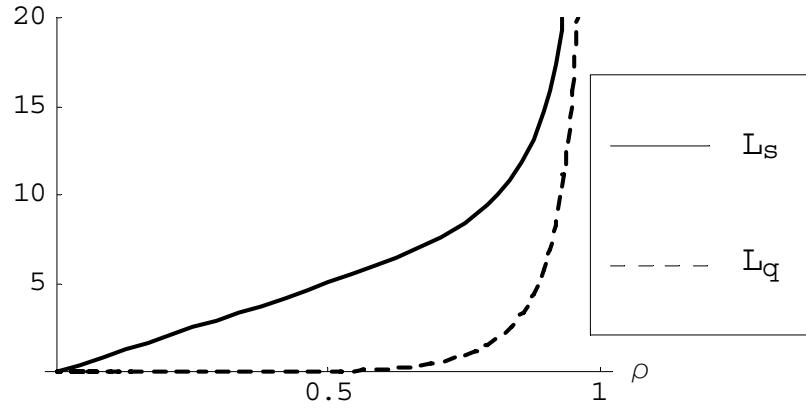


Figura 10: Número médio de clientes no sistema (L_s) e na fila (L_q) em função da taxa de ocupação (ρ), com $S = 10$

taxa de ocupação tende para 1. Quando a taxa de ocupação, no limite, é zero, o número de clientes no sistema é também 0, e consoante a taxa de ocupação aumenta o número de clientes no sistema aumenta, moderadamente no início e mais rapidamente quando ρ tende para 1.

Vamos agora determinar o tempo de espera no sistema, W_s . Usando (6) temos que

$$W_s = \frac{L_s}{\lambda}.$$

Substituindo L_s ,

$$\begin{aligned} W_s &= \frac{\frac{\lambda}{\mu} + \frac{\lambda}{S\mu} \frac{P_S}{\left(1 - \frac{\lambda}{S\mu}\right)^2}}{\lambda} \\ &= \frac{1}{\mu} + \frac{P_S}{S\mu} \frac{1}{\left(1 - \frac{\lambda}{S\mu}\right)^2}. \end{aligned}$$

Usando a taxa de ocupação,

$$W_s = \frac{1}{\mu} + \frac{P_S}{S\mu} \frac{1}{(1 - \rho)^2}.$$

Vamos supor que existem só 10 servidores para representar W_s no gráfico da Figura 11.

Observando a Figura 11, verificamos que independentemente da taxa de serviço, o tempo médio dos clientes no sistema cresce muito lentamente no início e tende para infinito quando a taxa de ocupação tende para 1. Quando a taxa de ocupação converge para 0 temos

$$\begin{aligned} \lim_{\rho \rightarrow 0} W_s &= \lim_{\rho \rightarrow 0} \left[\frac{1}{\mu} + \frac{\frac{S^n}{n!} \rho^n}{S\mu (1 - \rho)^2 \left(\sum_{0 \leq n \leq S-1} \frac{S^n}{n!} \rho^n + \frac{S^S}{S!} \rho^S \left(\frac{1}{1 - \rho} \right) \right)} \right] \\ &= \frac{1}{\mu}. \end{aligned}$$

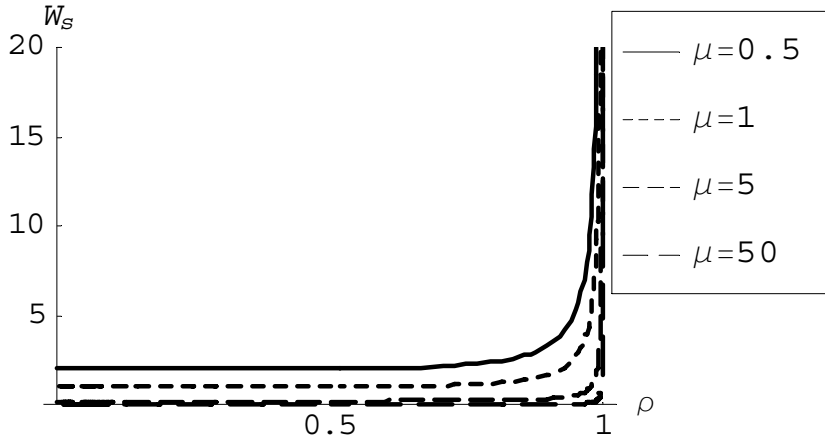


Figura 11: Tempo médio de espera no sistema (W_s) em função da taxa de ocupação (ρ), com $S = 10$

Ainda nos falta calcular o tempo de espera na fila, W_q . Usando (7), resulta que

$$W_q = \frac{L_q}{\lambda}.$$

Substituindo L_q pela expressão obtida em (18) obtemos

$$\begin{aligned} W_q &= \frac{\frac{\lambda}{S\mu} \frac{P_S}{\left(1 - \frac{\lambda}{S\mu}\right)^2}}{\lambda} \\ &= \frac{P_S}{S\mu} \frac{1}{\left(1 - \frac{\lambda}{S\mu}\right)^2}. \end{aligned}$$

que é equivalente a

$$W_q = \frac{P_S}{S\mu} \frac{1}{(1 - \rho)^2}.$$

Supondo que temos 10 servidores pode-se representar graficamente W_q na Figura 12.

Podemos verificar que a situação que ocorre na Figura 12 não difere muito da situação da Figura 11.

Quando um cliente entra no sistema se existir $M = m < S$ clientes à sua frente, o cliente será servido de imediato, e o tempo esperado no sistema é $\frac{1}{\mu}$. Caso $m \geq S$ ele terá de esperar até que o cliente $(m - S + 1)$ - ésimo termine de ser servido.

A probabilidade de que o cliente não tenha de esperar é então

$$\begin{aligned} P[T_q = 0] &= P[M < S] = \sum_{n=0}^{S-1} P_n = \sum_{n=0}^S P_n - P_S \\ &= 1 - \left(\frac{\lambda}{S\mu}\right) \frac{P_S}{1 - \frac{\lambda}{S\mu}} - P_S = 1 - P_S \left(\frac{\lambda}{S\mu} \frac{1}{1 - \frac{\lambda}{S\mu}} + 1\right) \\ &= 1 - \frac{P_S}{1 - \frac{\lambda}{S\mu}}, \end{aligned}$$

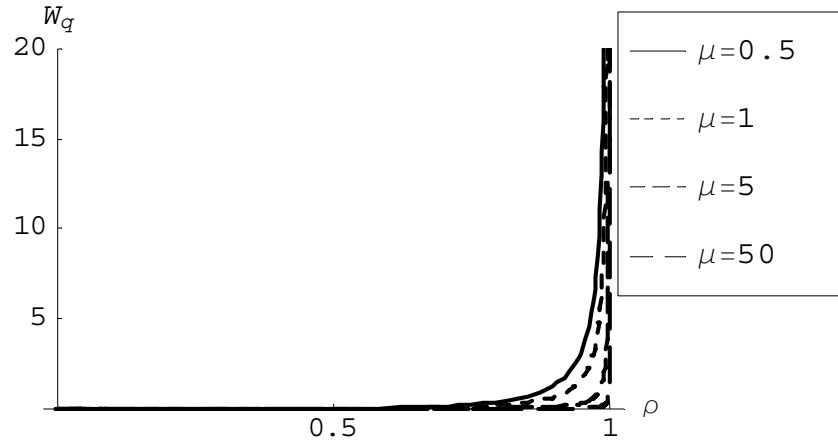


Figura 12: Tempo médio de espera na fila (W_q) em função da taxa de ocupação (ρ), com $S = 10$

que é equivalente a

$$P[T_q = 0] = 1 - \frac{P_S}{1 - \rho}.$$

Supondo que existem $S = 10$ servidores, ilustramos na Figura 13, $P[T_q = 0]$ como função de ρ .

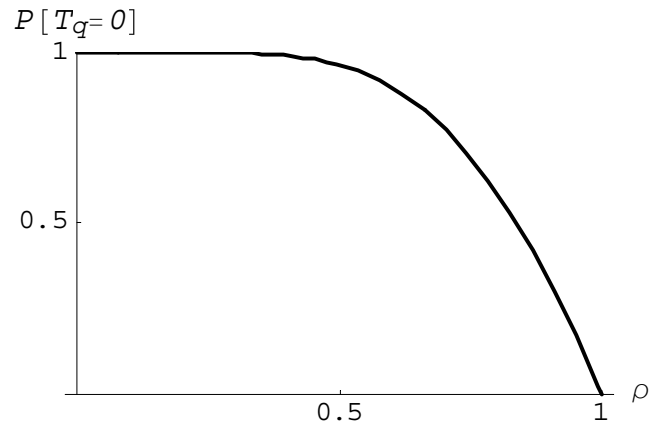


Figura 13: Probabilidade de o tempo de espera na fila ser zero, $P[T_q = 0]$, em função da taxa de ocupação, ρ , com $S = 10$

Analisando a Figura 13, observamos que a probabilidade do tempo de espera na fila dos clientes começa por 1, no caso limite $\rho = 0$, e praticamente se mantém neste valor até, sensivelmente, $\rho = \frac{1}{2}$, decrescendo mais rapidamente para 0, à medida que ρ cresce para 1.

Temos,

$$\begin{aligned}
 \lim_{\rho \rightarrow 1^-} P[T_q = 0] &= \lim_{\rho \rightarrow 1^-} \left(1 - \frac{\frac{S^S}{S!} \rho^S}{(1-\rho) \left(\sum_{0 \leq n \leq S-1} \frac{S^n}{n!} \rho^n + \frac{S^S}{S!} \rho^S \left(\frac{1}{1-\rho} \right) \right)} \right) \\
 &= \lim_{\rho \rightarrow 1^-} \left(1 - \frac{\frac{S^S}{S!} \rho^S}{(1-\rho) \sum_{0 \leq n \leq S-1} \frac{S^n}{n!} \rho^n + \frac{S^S}{S!} \rho^S} \right) = 1 - 1 \\
 &= 0.
 \end{aligned}$$

Se todos os servidores estiverem ocupados, os intervalos entre as sucessivas partidas são independentes com distribuição exponencial de parâmetro $S\mu$ pelo que o tempo total até à partida $m - S + 1$ tem distribuição Gama $(m - S + 1, S\mu)$, logo, numa situação de equilíbrio, temos,

$$\begin{aligned}
 P[0 < T_q \leq x] &= \sum_{m=S}^{+\infty} P[0 < T_q \leq x | M = m] P[M = m] \\
 &= \sum_{m=S}^{+\infty} \left[\int_0^x (S\mu) e^{-S\mu t} \frac{(S\mu t)^{m-S}}{(m-S)!} dt \right] P_m \\
 &= \int_0^x (S\mu) e^{-S\mu t} \left[\sum_{m=S}^{+\infty} \frac{(S\mu t)^{m-S}}{(m-S)!} \frac{1}{S!} \left(\frac{\lambda}{S\mu} \right)^{m-S} \left(\frac{\lambda}{\mu} \right)^S P_0 \right] dt \\
 &= \int_0^x (S\mu) e^{-S\mu t} \left[\frac{1}{S!} \left(\frac{\lambda}{\mu} \right)^S P_0 \sum_{m=S}^{+\infty} \frac{(t\lambda)^{m-S}}{(m-S)!} \right] dt = \int_0^x (S\mu) e^{-S\mu t} P_S e^{t\lambda} dt \\
 &= S\mu P_S \int_0^x e^{-(S\mu-\lambda)t} dt = S\mu P_S \left[\frac{e^{-(S\mu-\lambda)t}}{-(S\mu-\lambda)} \right]_{t=0}^x = \frac{P_S}{1 - \frac{\lambda}{S\mu}} [1 - e^{-(S\mu-\lambda)x}] \\
 &= \frac{P_S}{1-\rho} [1 - e^{-S\mu(1-\rho)x}].
 \end{aligned}$$

O tempo de espera tem, para valores superiores a zero “função densidade” dada por

$$S\mu P_S e^{-(S\mu-\lambda)x}.$$

Se $t \geq 0$

$$\begin{aligned}
 P[T_q \leq t] &= P(T_q = 0) + P[0 < T_q \leq t] \\
 &= 1 - \frac{P_S}{1-\rho} + \frac{P_S}{1-\rho} [1 - e^{-S\mu(1-\rho)t}] = 1 - \frac{P_S}{(1-\rho)} e^{-(S\mu-\lambda)t} \\
 &= 1 - \frac{P_S}{(1-\rho)} e^{-S\mu(1-\rho)t}.
 \end{aligned}$$

Resulta assim que o tempo de espera na fila exceder t é

$$P[T_q > t] = 1 - P[T_q \leq t] = \frac{P_S}{(1-\rho)} e^{-(S\mu-\lambda)t}.$$

O quadro que se segue é o quadro resumo do modelo $M/M/S$.

Probabilidade de ocorrência do estado 0	$P_0 = \frac{1}{\sum_{0 \leq n \leq S-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{S!} \left(\frac{\lambda}{\mu}\right)^S \left(\frac{1}{1-\rho}\right)}$
Probabilidade de ocorrência do estado n	$P_n = \begin{cases} \frac{S^n}{n!} \rho^n P_0 & \text{se } n = 0, 1, \dots, S \\ \frac{S!}{S!} \rho^n P_0 & \text{se } n > S \end{cases}$
Número médio de clientes no sistema	$L_s = \frac{\lambda}{\mu} + \rho \frac{P_s}{(1-\rho)^2}$
Comprimento médio da fila	$L_q = \rho \frac{P_s}{(1-\rho)^2}$
Tempo médio de espera no sistema	$W_s = \frac{1}{\mu} + \frac{P_s}{S\mu} \frac{1}{(1-\rho)^2}$
Tempo médio de espera na fila	$W_q = \frac{P_s}{S\mu} \frac{1}{(1-\rho)^2}$
Probabilidade de o tempo de espera na fila ser zero	$P(T_q = 0) = 1 - \frac{P_s}{1-\rho}$
Probabilidade de o tempo de espera na fila exceder t	$P(T_q > t) = \frac{P_s}{(1-\rho)} e^{-(S\mu - \lambda)t}$
Taxa de ocupação	$\rho = \frac{\lambda}{S\mu}$

Quadro 2: Sumário das características do modelo $M/M/S$

4.3 Modelo $M/M/\infty$

Neste modelo o número de servidores é ilimitado, a distribuição dos tempos entre as chegadas e dos tempos de atendimento é exponencial, a capacidade do sistema e da população são infinitas e a disciplina corresponde a quem entra primeiro no sistema ser o primeiro a ser atendido e a sair. Um exemplo possível de aplicação deste modelo é o sistema de linhas telefónicas (actualmente).

Os clientes ao chegarem ao sistema são servidos de imediato. Se a taxa de partida de um único cliente for μ , a taxa de partida quando existem k clientes é $k\mu$, e a taxa de ocupação é $\rho = \frac{\lambda}{\mu}$. Este sistema pode ser representado através da Figura 14.

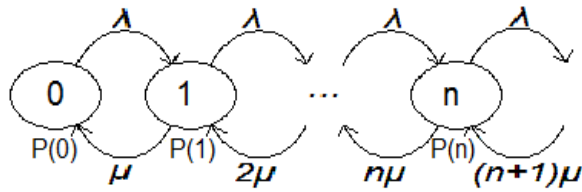


Figura 14: Modelo $M/M/\infty$

Neste caso a distribuição limite, tendo em conta que

$$\theta_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} = \frac{\lambda^n}{n! \mu^n}.$$

é dada por (cf. Subsecção 2.4.1)

$$\begin{aligned}
P_n &= \lim_{t \rightarrow +\infty} P[X(t) = n] = \frac{\theta_n}{\sum_{n \geq 0} \theta_n} = \frac{\frac{\lambda^n}{n! \mu^n}}{\sum_{n \geq 0} \frac{\lambda^n}{n! \mu^n}} \\
&= \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} e^{-\frac{\lambda}{\mu}} = \frac{\rho^n}{n!} e^{-\rho}, (\forall \rho > 0),
\end{aligned}$$

donde se conclui que a distribuição limite corresponde à distribuição de Poisson de parâmetro ρ , e pode ser representado pela Figura 15.

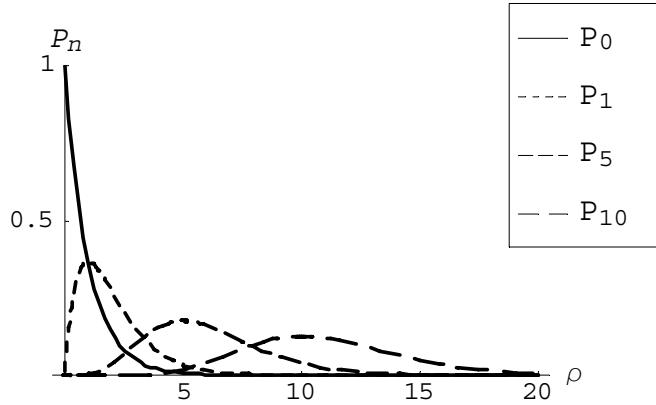


Figura 15: Probabilidade de existirem n clientes no sistema (P_n) em função da taxa de ocupação (ρ)

Podemos observar na Figura 15 que a probabilidade de existirem zero clientes é 1 quando a taxa de ocupação é zero, e tende a diminuir consoante aumenta a taxa de ocupação. Quanto à probabilidade de existir 1 cliente no sistema começa em zero quando a taxa de ocupação é zero, e tende a aumentar, mas não chegando ao valor 0.5, e voltando depois e decrescer até zero. Em relação a probabilidade de existirem 2 ou mais clientes no sistema a situação é análoga à anterior.

O número médio de clientes no sistema é simplesmente igual à taxa de ocupação

$$L_s = \rho.$$

(cf. Figura 16),

O tempo esperado que cada cliente passe no sistema é o tempo que ele demora a ser atendido, ou seja,

$$W_s = \frac{1}{\mu},$$

função que é representada na Figura 17.

O tempo de espera no sistema é inversamente proporcional à taxa de serviço, ou seja, enquanto que a taxa de serviço aumenta, diminui o tempo de espera no serviço no serviço, como já era de esperar.

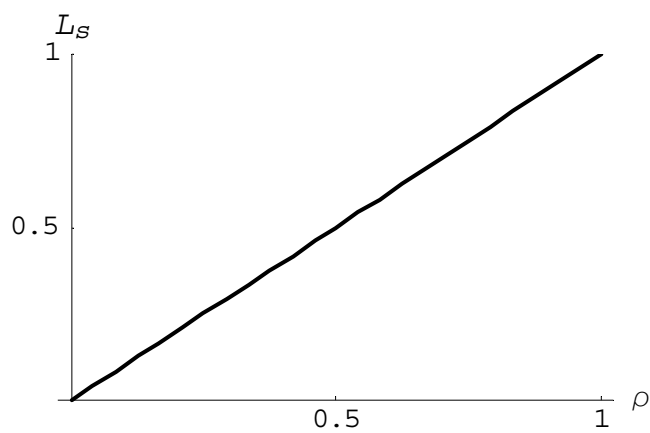


Figura 16: Número médio de clientes no sistema (L_s) em função da taxa de ocupação (ρ)

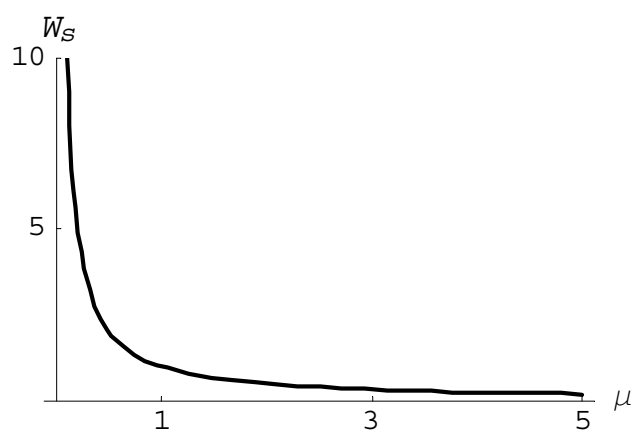


Figura 17: Tempo médio de espera no sistema (W_s) em função da taxa de serviço (μ)

Relativamente ao número médio de clientes na fila e ao tempo esperado na fila, temos

$$L_q = W_q = 0.$$

Podemos resumir as características do modelo $M/M/\infty$, no seguinte quadro;

Probabilidade de ocorrência do estado 0	$P_0 = e^{-\rho}$
Probabilidade de ocorrência do estado n	$P_n = \frac{\rho^n}{n!} e^{-\rho}$
Número médio de clientes no sistema	$L_s = \rho$
Comprimento médio da fila	$L_q = 0$
Tempo médio de espera no sistema	$W_s = \frac{1}{\mu}$
Tempo médio de espera na fila	$W_q = 0$
Taxa de ocupação	$\rho = \frac{\lambda}{\mu}$

Quadro 3: Sumário das características do modelo $M/M/\infty$

4.4 Modelo $M/M/1/K$

Neste modelo o sistema tem uma distribuição dos tempos entre as chegadas e dos tempos de atendimento exponencial, contém um só servidor, a capacidade do sistema é K , a população é infinita e a disciplina corresponde a quem entra primeiro no sistema ser o primeiro a ser atendido e a sair. Este sistema pode também ser representado por $M/M/1/K/\infty/FIFO$ ou por $M/M/1/K/\infty/FCFS$. Um exemplo deste modelo é de uma estação de serviço, com um número limitado para K veículos e com uma única bomba de abastecimento.

Recorde-se que (cf Subsecção 3.1.4) a capacidade do sistema ser K significa que a probabilidade de o sistema estar num estado $n \geq K+1$ é nula, pois o sistema não aceita ninguém se já tiver K clientes.

Neste modelo a taxa de chegada, λ_n , depende do estado n do sistema,

$$\lambda_n = \begin{cases} \lambda, & \text{se } n = 0, 1, 2, \dots, K-1 \\ 0, & \text{se } n \geq K \end{cases},$$

bem como a taxa de saída,

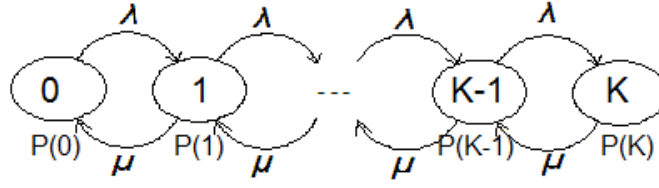
$$\mu_n = \begin{cases} \mu, & \text{se } n = 0, 1, 2, \dots, K \\ 0, & \text{se } n > K \end{cases}.$$

As transições entre os vários estados deste modelo podem ser representadas através da Figura 18.

A taxa média de entradas no sistema λ' , é dada por

$$\lambda' = \sum_{n=0}^{+\infty} \lambda_n P_n = \sum_{n=0}^{K-1} \lambda P_n = \lambda \sum_{n=0}^{K-1} P_n = \lambda (1 - P_K),$$

onde P_K é a probabilidade de desistência por falta de capacidade do sistema para atender clientes que chegam quando o sistema está cheio.

Figura 18: Modelo $M/M/1/K$

Como foi referido na Secção 3.3.1, as relações fundamentais entre L_q, L_s, W_q e W_s mantêm-se verdadeiras substituindo apenas λ por λ' , a taxa de ocupação é $\frac{\lambda'}{\mu}$ e a taxa de pressão é designado por $\rho = \frac{\lambda}{\mu}$.

Como veremos, o sistema pode estar em equilíbrio para valores de ρ superiores a 1. No entanto, nesse caso, haverá um número grande de clientes que chegam e não são atendidos.

Muitas deduções do modelo $M/M/1$ são válidas, com as devidas adaptações para ter em conta o limite de clientes no sistema. Adaptando a expressão (12) temos

$$P_n = \begin{cases} \rho^n P_0, & \text{se } n = 0, 1, \dots, K \\ 0, & \text{se } n > K \end{cases} \quad (19)$$

Para podermos calcular P_0 temos que recorrer ao facto de

$$1 = P_0 + \dots + P_K.$$

Usando (19) temos

$$\begin{aligned} 1 &= P_0 + \dots + P_K = P_0 + \dots + \rho^K P_0 \\ &= P_0 \sum_{n=0}^K \rho^n = P_0 \left(\frac{1 - \rho^{K+1}}{1 - \rho} \right), \end{aligned}$$

se $\rho \neq 1$. Mais

$$1 = P_0 \left(\frac{1 - \rho^{K+1}}{1 - \rho} \right) \iff P_0 = \frac{1 - \rho}{1 - \rho^{K+1}},$$

se $\rho \neq 1$.

Quando $\rho = 1$, temos

$$P_n = P_0,$$

e

$$\begin{aligned} 1 &= P_0 + \dots + P_K = P_0 + \dots + 1^K P_0 = P_0 (K + 1) \\ \Rightarrow P_0 &= \frac{1}{K + 1}, \end{aligned}$$

ou seja,

$$P_0 = \begin{cases} \frac{1 - \rho}{1 - \rho^{K+1}}, & \text{se } \rho \neq 1 \\ \frac{1}{K+1}, & \text{se } \rho = 1 \end{cases} \quad (20)$$

Podemos reescrever a expressão (19):

$$P_n = \begin{cases} \rho^n \left(\frac{1 - \rho}{1 - \rho^{K+1}} \right), & \text{se } \rho \neq 1, n = 0, 1, \dots, K \\ \frac{1}{K+1}, & \text{se } \rho = 1, n = 0, 1, \dots, K \\ 0, & \text{se } n > K \end{cases} \quad (21)$$

Supondo que $K = 20$, representamos P_n na Figura 19, em função da taxa de pressão, ρ , fazendo ρ variar de 0 a 1.5 e considerando $n = 0, 1, 5$ e 20. Na Figura 20 fazemos ρ variar de 0 a 20 e tomando $n = 20$.

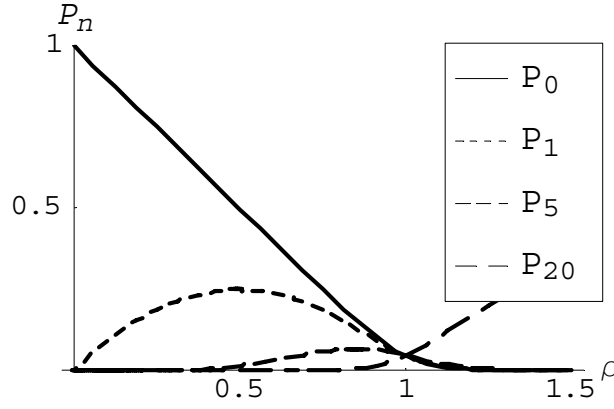


Figura 19: Probabilidade de existirem n clientes no sistema (P_n) em função da taxa de pressão (ρ), com $K = 20$

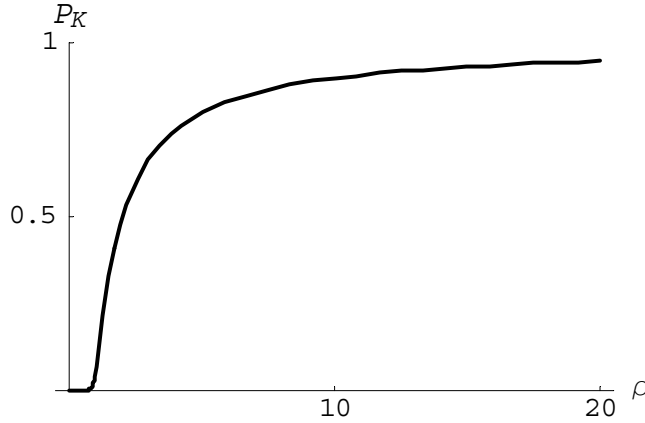


Figura 20: Probabilidade de existirem K clientes no sistema (P_K) em função da taxa de pressão (ρ), com $K = 20$

Verificamos na Figura 19 que a probabilidade de existirem zero clientes no sistema é 1 quando a taxa de pressão é zero e diminui consoante aumenta a taxa de pressão, atingindo o valor 0. Em relação às probabilidades de existir 1 cliente a $K - 1$ clientes começam em 0 aumentando um pouco e voltando a 0. E na Figura 20 mostra que a probabilidade de existirem K clientes no sistema converge para 1 quando ρ tende para o infinito, pois,

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} P_K &= \lim_{\rho \rightarrow +\infty} \left[\rho^K \left(\frac{1 - \rho}{1 - \rho^{K+1}} \right) \right] = \lim_{\rho \rightarrow +\infty} \left(\frac{\rho^K - \rho^{K+1}}{1 - \rho^{K+1}} \right) = \lim_{\rho \rightarrow +\infty} \left(\frac{\frac{\rho^K}{\rho^{K+1}} - 1}{\frac{1}{\rho^{K+1}} - 1} \right) \\ &= 1. \end{aligned}$$

Conhecendo a expressão para P_K (cf. (21)) podemos clarificar a expressão de λ' . Para $\rho \neq 1$,

$$\begin{aligned}\lambda' &= \lambda(1 - P_K) = \lambda \left(1 - \rho^K \left(\frac{1 - \rho}{1 - \rho^{K+1}} \right) \right) = \lambda \left(\frac{1 - \rho^{K+1}}{1 - \rho^{K+1}} - \frac{\rho^K - \rho^{K+1}}{1 - \rho^{K+1}} \right) \\ &= \lambda \left(\frac{1 - \rho^K}{1 - \rho^{K+1}} \right).\end{aligned}$$

Quando $\rho = 1$,

$$\lambda' = \lambda \left(1 - \frac{1}{K+1} \right) = \lambda \left(\frac{K}{K+1} \right).$$

Ou seja,

$$\lambda' = \begin{cases} \lambda \left(\frac{1 - \rho^K}{1 - \rho^{K+1}} \right), & \text{se } \rho \neq 1 \\ \lambda \left(\frac{K}{K+1} \right), & \text{se } \rho = 1 \end{cases}.$$

Vamos agora calcular o valor médio de clientes no sistema,

$$\begin{aligned}L_s &= \sum_{n=0}^K nP_n = \sum_{n=1}^K n\rho^n \left(\frac{1 - \rho}{1 - \rho^{K+1}} \right) = \frac{1 - \rho}{1 - \rho^{K+1}} \sum_{n=1}^K n\rho^n = \frac{1 - \rho}{1 - \rho^{K+1}} \rho \sum_{n=1}^K n\rho^{n-1} \\ &= \frac{1 - \rho}{1 - \rho^{K+1}} \rho \sum_{n=1}^K \frac{d\rho^n}{d\rho} = \frac{1 - \rho}{1 - \rho^{K+1}} \rho \frac{d}{d\rho} \sum_{n=0}^K \rho^n = \frac{1 - \rho}{1 - \rho^{K+1}} \rho \frac{d}{d\rho} \left(\frac{1 - \rho^{K+1}}{1 - \rho} \right) \\ &= \frac{1 - \rho}{1 - \rho^{K+1}} \rho \left(\frac{-(K+1)\rho^K(1 - \rho) + (1 - \rho^{K+1})}{(1 - \rho)^2} \right) \\ &= \frac{\rho}{1 - \rho^{K+1}} \left(-(K+1)\rho^K + \frac{1 - \rho^{K+1}}{1 - \rho} \right) \\ &= \frac{\rho}{1 - \rho} - \frac{(K+1)\rho^{K+1}}{1 - \rho^{K+1}},\end{aligned}$$

quando $\rho \neq 1$.

Quando $\rho = 1$,

$$L_s = \sum_{n=0}^K nP_n = \frac{1}{K+1} (1 + 2 + \dots + K) = \frac{K(K+1)}{2(K+1)} = \frac{K}{2}.$$

Logo,

$$L_s = \begin{cases} \frac{\rho}{1 - \rho} - \frac{(K+1)\rho^{K+1}}{1 - \rho^{K+1}}, & \text{se } \rho \neq 1 \\ \frac{K}{2}, & \text{se } \rho = 1 \end{cases}.$$

Vamos calcular o valor médio de clientes na fila usando a expressão, tendo em conta (9) que λ depende do estado do sistema e que deve ser substituído por λ' , isto é,

$$L_q = L_s - \frac{\lambda'}{\mu}.$$

Para $\rho \neq 1$,

$$\begin{aligned}
 L_q &= \frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}} - \frac{\lambda \left(\frac{1-\rho^K}{1-\rho^{K+1}} \right)}{\mu} = \frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}} - \rho \left(\frac{1-\rho^K}{1-\rho^{K+1}} \right) \\
 &= \frac{\rho}{1-\rho} + \frac{-K\rho^{K+1} - \rho^{K+1}}{1-\rho^{K+1}} + \frac{-\rho + \rho^{K+1}}{1-\rho^{K+1}} = \frac{\rho}{1-\rho} + \frac{-K\rho^{K+1} - \rho^{K+1} - \rho + \rho^{K+1}}{1-\rho^{K+1}} \\
 &= \frac{\rho}{1-\rho} - \frac{K\rho^{K+1} + \rho}{1-\rho^{K+1}}.
 \end{aligned}$$

Para $\rho = 1$,

$$\begin{aligned}
 L_q &= \frac{K}{2} - \frac{\lambda \left(\frac{K}{K+1} \right)}{\mu} = \frac{K}{2} - \rho \left(\frac{K}{K+1} \right) = \frac{K}{2} - \frac{K}{K+1} \\
 &= \frac{K(K+1) - 2K}{2(K+1)} = \frac{K^2 + K - 2K}{2(K+1)} \\
 &= \frac{K(K-1)}{2(K+1)}.
 \end{aligned}$$

Conclui-se que

$$L_q = \begin{cases} \frac{\rho}{1-\rho} - \frac{K\rho^{K+1} + \rho}{1-\rho^{K+1}}, & \text{se } \rho \neq 1 \\ \frac{K(K-1)}{2(K+1)}, & \text{se } \rho = 1 \end{cases}.$$

Supondo novamente que $K = 20$, o gráfico da Figura 21 representa L_s e L_q como funções de ρ .

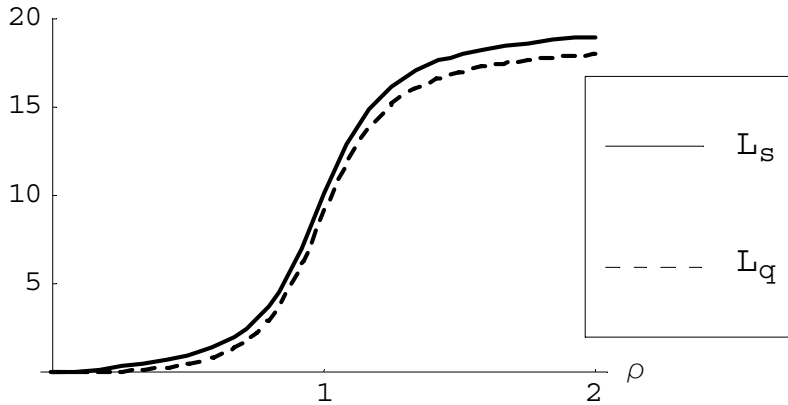


Figura 21: Número médio de clientes no sistema (L_s) e na fila (L_q) em função da taxa de pressão (ρ), com $K = 20$

Analisando a Figura 21 verificamos que, quando na situação limite da taxa de pressão ser zero, o número médio de clientes no sistema e na fila também são 0. Estes valores aumentam ao longo do crescimento da taxa de pressão, nunca ultrapassando K . Quando ρ converge

para $+\infty$, L_s converge para K e L_q converge para $K - 1$. De facto,

$$\begin{aligned}\lim_{\rho \rightarrow +\infty} L_s &= \lim_{\rho \rightarrow +\infty} \left[\frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{1}{\frac{1}{\rho} - 1} - \frac{K+1}{\frac{1}{\rho^{K+1}} - 1} \right] \\ &= \frac{1}{-1} - \frac{K+1}{-1} = -1 + K + 1 \\ &= K\end{aligned}$$

e

$$\begin{aligned}\lim_{\rho \rightarrow +\infty} L_q &= \lim_{\rho \rightarrow +\infty} \left[\frac{\rho}{1-\rho} - \frac{K\rho^{K+1} + \rho}{1-\rho^{K+1}} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{1}{\frac{1}{\rho} - 1} - \frac{K + \frac{1}{\rho^K}}{\frac{1}{\rho^{K+1}} - 1} \right] \\ &= \frac{1}{-1} - \frac{K}{-1} \\ &= K - 1.\end{aligned}$$

Usando a Lei de Little, com λ substituído por λ' , vamos calcular o tempo de espera na fila. Se $\rho \neq 1$,

$$\begin{aligned}W_q &= \frac{L_q}{\lambda'} = \frac{\frac{\rho}{1-\rho} - \frac{K\rho^{K+1} + \rho}{1-\rho^{K+1}}}{\lambda \left(\frac{1-\rho^K}{1-\rho^{K+1}} \right)} = \frac{\frac{\rho - \rho^{K+2}}{(1-\rho)(1-\rho^{K+1})} - \frac{K\rho^{K+1} + \rho - K\rho^{K+2} - \rho^2}{(1-\rho)(1-\rho^{K+1})}}{\lambda \left(\frac{1-\rho^K}{1-\rho^{K+1}} \right)} \\ &= \frac{\rho - \rho^{K+2} - (K\rho^{K+1} + \rho - K\rho^{K+2} - \rho^2)}{\lambda(1-\rho)(1-\rho^K)} \\ &= \frac{\rho - \rho^{K+2} - K\rho^{K+1} - \rho + K\rho^{K+2} + \rho^2}{\lambda(1-\rho)(1-\rho^K)} \\ &= \frac{-\rho^{K+2} - K\rho^{K+1} + K\rho^{K+2} + \rho^2}{\lambda(1-\rho)(1-\rho^K)} = \frac{K\rho^{K+1}(\rho - 1) + \rho^2(1 - \rho^K)}{\lambda(1-\rho)(1-\rho^K)} \\ &= -\frac{K\rho^{K+1}(1-\rho)}{\lambda(1-\rho)(1-\rho^K)} + \frac{\rho^2(1-\rho^K)}{\lambda(1-\rho)(1-\rho^K)} = \frac{\rho^2}{\lambda(1-\rho)} - \frac{K\rho^{K+1}}{\lambda(1-\rho^K)} \\ &= \frac{\rho}{\mu(1-\rho)} - \frac{K\rho^K}{\mu(1-\rho^K)}.\end{aligned}$$

Quando $\rho = 1$,

$$\begin{aligned}W_q &= \frac{L_q}{\lambda'} = \frac{\frac{K(K-1)}{2(K+1)}}{\lambda \left(\frac{K}{K+1} \right)} = \frac{K(K-1)}{2\lambda K} \\ &= \frac{K-1}{2\rho\mu}.\end{aligned}$$

Logo,

$$W_q = \begin{cases} \frac{\rho}{\mu(1-\rho)} - \frac{K\rho^K}{\mu(1-\rho^K)}, & \text{se } \rho \neq 1 \\ \frac{K-1}{2\rho\mu}, & \text{se } \rho = 1 \end{cases}.$$

Na Figura 22, supondo que $K = 20$, representamos W_q como função de ρ para vários valores de μ ($\mu = 0.5, 1, 5, 50$).

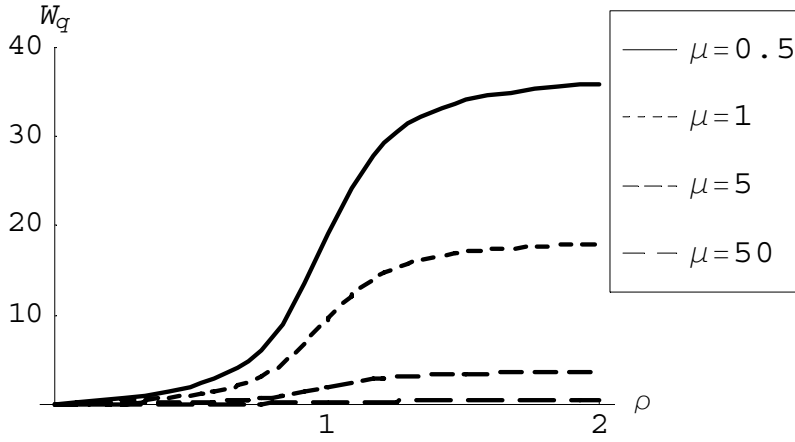


Figura 22: Tempo médio de espera na fila (W_q) em função da taxa de pressão (ρ), com $K = 20$

Podemos concluir, analisando a Figura 22, que independentemente da taxa de serviço, o tempo médio de um cliente na fila começa em zero, aumenta com o aumento da taxa de pressão e depois “estabiliza”. De facto,

$$\begin{aligned}
 \lim_{\rho \rightarrow +\infty} W_q &= \lim_{\rho \rightarrow +\infty} \left[\frac{\rho}{\mu(1-\rho)} - \frac{K\rho^K}{\mu(1-\rho^K)} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{1}{\frac{\rho}{\mu} - \mu} - \frac{K}{\frac{\rho^K}{\mu} - \mu} \right] \\
 &= \frac{K-1}{\mu}.
 \end{aligned} \tag{22}$$

Em relação à evolução de W_q como função da taxa de serviço, podemos dizer que quanto maior é a taxa de serviço menos tempo em média o cliente fica na fila à espera, naturalmente.

Falta-nos calcular o tempo de espera no sistema. Ora,

$$\begin{aligned}
 W_s &= \frac{L_s}{\lambda'} = \frac{\frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}}}{\lambda \left(\frac{1-\rho^K}{1-\rho^{K+1}} \right)} = \frac{\frac{\rho - \rho^{K+2}}{(1-\rho)(1-\rho^{K+1})} - \frac{(K+1)\rho^{K+1} - (K+1)\rho^{K+2}}{(1-\rho)(1-\rho^{K+1})}}{\lambda \left(\frac{1-\rho^K}{1-\rho^{K+1}} \right)} \\
 &= \frac{\rho - \rho^{K+2} - ((K+1)\rho^{K+1} - (K+1)\rho^{K+2})}{\lambda(1-\rho)(1-\rho^K)} \\
 &= \frac{\rho - \rho^{K+2} - K\rho^{K+1} - \rho^{K+1} + K\rho^{K+2} + \rho^{K+2}}{\lambda(1-\rho)(1-\rho^K)} \\
 &= \frac{\rho - K\rho^{K+1} - \rho^{K+1} + K\rho^{K+2}}{\lambda(1-\rho)(1-\rho^K)} = \frac{K\rho^{K+1}(\rho-1) + \rho(1-\rho^K)}{\lambda(1-\rho)(1-\rho^K)} \\
 &= \frac{\rho(1-\rho^K)}{\lambda(1-\rho)(1-\rho^K)} - \frac{K\rho^{K+1}(1-\rho)}{\lambda(1-\rho)(1-\rho^K)} = \frac{\rho}{\lambda(1-\rho)} - \frac{K\rho^{K+1}}{\lambda(1-\rho^K)} \\
 &= \frac{1}{\mu(1-\rho)} - \frac{K\rho^K}{\mu(1-\rho^K)}
 \end{aligned}$$

se $\rho \neq 1$.

E quando $\rho = 1$,

$$\begin{aligned} W_s &= \frac{L_s}{\lambda'} = \frac{\frac{K}{2}}{\lambda \left(\frac{K}{K+1} \right)} = \frac{K(K+1)}{2\lambda K} \\ &= \frac{K+1}{2\rho\mu}. \end{aligned}$$

Conclui-se que

$$W_s = \begin{cases} \frac{1}{\mu(1-\rho)} - \frac{K\rho^K}{\mu(1-\rho^K)}, & \text{se } \rho \neq 1 \\ \frac{K+1}{2\rho\mu}, & \text{se } \rho = 1 \end{cases}.$$

Representamos W_s no gráfico da Figura 23 como função de ρ , considerando $K = 20$, e $\mu = 0.5, 1, 5$ e 50 .

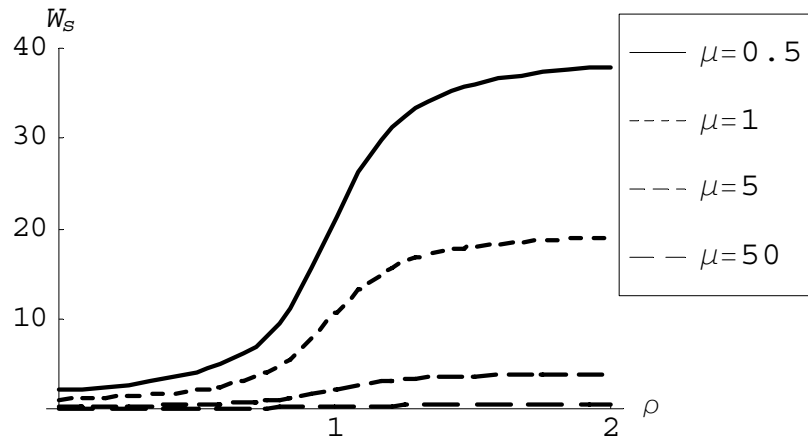


Figura 23: Tempo médio de espera no sistema (W_s) em função da taxa de pressão (ρ), com $K = 20$

O gráfico da Figura 23, é análogo ao anterior, o da Figura 22, como esperado, pois $W_s = W_q + \frac{1}{\mu}$. No limite, recordando (22), temos

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} W_s &= \lim_{\rho \rightarrow +\infty} W_q + \frac{1}{\mu} = \frac{K-1}{\mu} + \frac{1}{\mu} \\ &= \frac{K}{\mu}. \end{aligned}$$

A probabilidade de não se esperar tempo nenhum na fila, $P[T_q = 0]$, é igual à probabilidade de haver 0 ocorrências, ou seja,

$$P[T_q = 0] = P_0 = \begin{cases} \frac{1-\rho}{1-\rho^{K+1}}, & \text{se } \rho \neq 1 \\ \frac{1}{K+1}, & \text{se } \rho = 1 \end{cases}.$$

Na Figura 24 representamos graficamente $P[T_q = 0]$ supondo que $K = 20$.

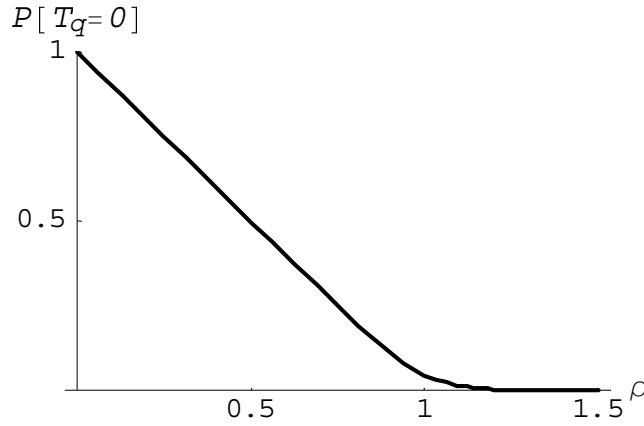


Figura 24: Probabilidade de o tempo de espera na fila ser zero, $P[T_q = 0]$, em função da taxa de pressão, ρ , com $K = 20$

Na Figura 24, observamos que a probabilidade de um cliente não ter de esperar na fila, começa em 1, quando a taxa de pressão toma o valor limite 0 e essa probabilidade converge para 0, quando a taxa de pressão converge para $+\infty$. De facto, de (20),

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} P_0 &= \lim_{\rho \rightarrow +\infty} \left[\frac{1 - \rho}{1 - \rho^{K+1}} \right] \\ &= 0. \end{aligned}$$

O quadro que se segue é o quadro resumo do modelo $M/M/1/K$.

Probabilidade de ocorrência do estado 0	$P_0 = \begin{cases} \frac{1-\rho}{1-\rho^{K+1}}, & \rho \neq 1 \\ \frac{1}{K+1}, & \rho = 1 \end{cases}$
Probabilidade de ocorrência do estado n	$P_n = \begin{cases} \rho^n \left(\frac{1-\rho}{1-\rho^{K+1}} \right), & \rho \neq 1, n = 0, 1, \dots, K \\ \frac{1}{K+1}, & \rho = 1, n = 0, 1, \dots, K \\ 0, & n > K \end{cases}$
Número médio de clientes no sistema	$L_s = \begin{cases} \frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}}, & \rho \neq 1 \\ \frac{K}{2}, & \rho = 1 \end{cases}$
Comprimento médio da fila	$L_q = \begin{cases} \frac{\rho}{1-\rho} - \frac{K\rho^{K+1} + \rho}{1-\rho^{K+1}}, & \rho \neq 1 \\ \frac{K(K-1)}{2(K+1)}, & \rho = 1 \end{cases}$
Tempo médio de espera no sistema	$W_s = \begin{cases} \frac{1}{\mu(1-\rho)} - \frac{K\rho^K}{\mu(1-\rho^K)}, & \rho \neq 1 \\ \frac{K+1}{2\rho\mu}, & \rho = 1 \end{cases}$
Tempo médio de espera na fila	$W_q = \begin{cases} \frac{\rho}{\mu(1-\rho)} - \frac{K\rho^K}{\mu(1-\rho^K)}, & \rho \neq 1 \\ \frac{K-1}{2\rho\mu}, & \rho = 1 \end{cases}$
Probabilidade de o tempo de espera na fila ser zero	$P(T_q = 0) = P_0$
Taxa de pressão	$\rho = \frac{\lambda}{\mu}$

Quadro 4: Sumário das características do modelo $M/M/1/K$

4.5 Modelo $M/M/S/K$

Este modelo difere do apresentado na secção anterior por ter S servidores, enquanto que no anterior tínhamos apenas 1 servidor. Resumindo, neste modelo o sistema tem uma distribuição dos tempos entre as chegadas e dos tempos de atendimento exponencial, contém S servidores, a capacidade do sistema é K , a capacidade da população é infinita e a disciplina corresponde a quem entra primeiro no sistema ser o primeiro a ser atendido e a sair. Um exemplo deste modelo pode ser uma estação de serviço com capacidade de K carros e S bombas de abastecimento.

Este modelo corresponde ao modelo já estudado anteriormente $M/M/S$ com sistema de perda, pois ao chegar ao sistema o $(K + 1)$ -ésimo cliente, este cliente já não poderá entrar no sistema, pois o sistema estará lotado, e este cliente é perdido.

Neste caso vamos considerar que $S > 1$ (o caso $S = 1$ já foi estudado) e que $S < K$ (se $S \geq K$, o tempo de espera para os clientes que não se perdem é zero). A taxa de saída do sistema depende do estado do sistema

$$\mu_n = \begin{cases} n\mu, & \text{se } n = 0, 1, 2, \dots, S-1 \\ S\mu, & \text{se } n = S, \dots, K \end{cases},$$

tal como depende a taxa de chegada, λ_n

$$\lambda_n = \begin{cases} \lambda, & \text{se } n = 0, 1, 2, \dots, K-1 \\ 0, & \text{se } n \geq K \end{cases}.$$

Este sistema é representado na Figura 25.

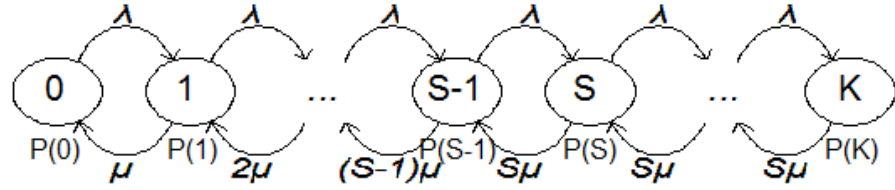


Figura 25: Modelo $M/M/S/K$

A taxa média de entradas no sistema λ' , é dada por

$$\lambda' = \sum_{n=0}^{K-1} \lambda P_n = \lambda \sum_{n=0}^{K-1} P_n = \lambda (1 - P_K),$$

onde P_K é a probabilidade de desistência por falta de capacidade do sistema para atender clientes que chegam quando o sistema está cheio, como no modelo anterior.

As relações fundamentais entre L_q , L_s , W_q e W_s obtemos substituindo apenas λ por λ' . A taxa de ocupação é $\frac{\lambda'}{S\mu}$ e a taxa de pressão é $\rho = \frac{\lambda'}{S\mu}$.

Existe alguma semelhança entre este modelo e o modelo $M/M/S$. Tendo em conta o limite de clientes no sistema, então temos que

$$P_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0, & \text{se } n = 0, 1, \dots, S-1 \\ \frac{1}{S! S^{n-S}} \left(\frac{\lambda}{\mu} \right)^n P_0, & \text{se } n = S, \dots, K \\ 0, & \text{se } n > K \end{cases},$$

que, tomando $\rho = \frac{\lambda}{S\mu}$, se pode escrever na forma

$$P_n = \begin{cases} \frac{(S\rho)^n}{n!} P_0, & \text{se } n = 0, 1, \dots, S-1 \\ \frac{S^S \rho^n}{S!} P_0, & \text{se } n = S, \dots, K \\ 0, & \text{se } n > K \end{cases} \quad (23)$$

E sendo assim temos que

$$\begin{aligned} \lambda' &= \lambda(1 - P_K) \\ &= \lambda \left(1 - \frac{S^S \rho^K}{S!} P_0 \right). \end{aligned}$$

Quando $\rho \neq 1$,

$$\begin{aligned} (P_0)^{-1} &= \sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \sum_{n=S}^K \frac{S^S \rho^n}{S!} = \sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \frac{S^S}{S!} \sum_{n=S}^K \rho^n \\ &= \sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \frac{S^S}{S!} \rho^S \frac{1 - \rho^{K-S+1}}{1 - \rho} \\ &= \sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \frac{(S\rho)^S}{S!} \left(\frac{1 - \rho^{K-S+1}}{1 - \rho} \right). \end{aligned}$$

Quando $\rho = 1$, isto é, quando $\lambda = S\mu$,

$$\begin{aligned} (P_0)^{-1} &= \sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \sum_{n=S}^K \frac{S^S \rho^n}{S!} = \sum_{n=0}^{S-1} \frac{S^n}{n!} + \frac{S^S}{S!} \sum_{n=S}^K 1 \\ &= \sum_{n=0}^{S-1} \frac{S^n}{n!} + \frac{S^S}{S!} (K - S + 1). \end{aligned}$$

Conclui-se que

$$P_0 = \begin{cases} \left[\sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \frac{(S\rho)^S}{S!} \left(\frac{1 - \rho^{K-S+1}}{1 - \rho} \right) \right]^{-1}, & \text{se } \rho \neq 1 \\ \left[\sum_{n=0}^{S-1} \frac{S^n}{n!} + \frac{S^S}{S!} (K - S + 1) \right]^{-1}, & \text{se } \rho = 1 \end{cases} \quad (24)$$

Podemos assim representar graficamente P_n , supondo, por exemplo, que $S = 10$ e $K = 30$. (ver Figura 26 quando ρ varia entre 0 e 1.5 e na Figura 27 quando ρ varia entre 0 e 20).

Podemos observar, na Figura 26, que a probabilidade de não existirem clientes no sistema é 1 quando a taxa de pressão é zero e depois diminui para 0 ao longo do aumento da taxa de pressão. Em relação à probabilidade do sistema ter n clientes, com $n \in \{1, \dots, K-1\}$, esta começa em 0, aumenta e volta a diminuir para 0, e quanto maior for a quantidade de clientes menor é esse aumento. Na Figura 27 observamos que a probabilidade de existir K clientes no sistema começa em 0 e aumenta, convergindo para 1, quando a taxa de pressão

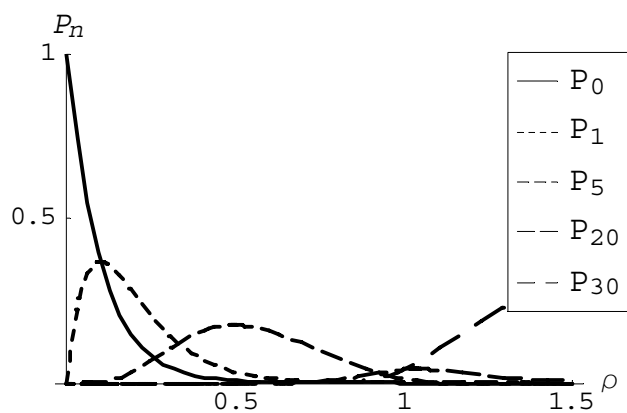


Figura 26: Probabilidade de existirem n clientes no sistema (P_n) em função da taxa de pressão (ρ), quando esta pertence ao intervalo $(0, 1.5)$, com $S = 10$ e $K = 30$

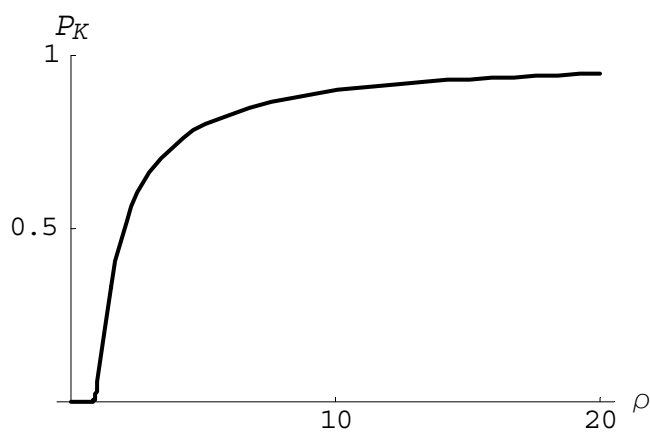


Figura 27: Probabilidade de existirem K clientes no sistema (P_K) em função da taxa de pressão (ρ), quando esta pertence ao intervalo $(0, 20)$, com $S = 10$ e $K = 30$

tende para infinito. De facto,

$$\begin{aligned}
 \lim_{\rho \rightarrow +\infty} P_K &= \lim_{\rho \rightarrow +\infty} \left[\frac{S^S \rho^K}{S! \left(\sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \frac{(S\rho)^S}{S!} \left(\frac{1-\rho^{K-S+1}}{1-\rho} \right) \right)} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{S^S \rho^K}{S! \frac{(S\rho)^S}{S!} \left(\frac{1-\rho^{K-S+1}}{1-\rho} \right)} \right] \\
 &= \lim_{\rho \rightarrow +\infty} \left[\frac{\rho^K}{\rho^S \left(\frac{1-\rho^{K-S+1}}{1-\rho} \right)} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{\rho^K (1-\rho)}{\rho^S - \rho^{K+1}} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{\frac{\rho^K}{\rho^{K+1}} - 1}{\frac{\rho^S}{\rho^{K+1}} - 1} \right] \\
 &= 1.
 \end{aligned}$$

Adaptando a derivação do número médio de clientes na fila do modelo $M/M/S$ temos

$$L_q = \sum_{n=S+1}^K (n-S) P_n.$$

Ora, para $\rho \neq 1$,

$$\begin{aligned}
 L_q &= \sum_{n=S+1}^K (n-S) \frac{S^S \rho^n}{S!} P_0 = P_0 \frac{S^S}{S!} \rho^{S+1} \sum_{n=S+1}^K (n-S) \rho^{n-S-1} \\
 &= P_0 \frac{S^S}{S!} \rho^{S+1} \sum_{j=1}^{K-S} j \rho^{j-1} = P_0 \frac{S^S}{S!} \rho^{S+1} \frac{d}{d\rho} \sum_{j=0}^{K-S} \rho^j = P_0 \frac{S^S}{S!} \rho^{S+1} \frac{d}{d\rho} \left(\frac{1-\rho^{K-S+1}}{1-\rho} \right) \\
 &= P_0 \frac{S^S}{S!} \rho^{S+1} \left(\frac{-(K-S+1) \rho^{K-S} (1-\rho) + 1 - \rho^{K-S+1}}{(1-\rho)^2} \right) \\
 &= P_0 \frac{S^S}{S!} \rho^{S+1} \left(\frac{1 - \rho^{K-S} - K \rho^{K-S} + K \rho^{K-S+1} + S \rho^{K-S} - S \rho^{K-S+1}}{(1-\rho)^2} \right) \\
 &= P_0 \frac{S^S}{S!} \rho^{S+1} \left(\frac{1 - \rho^{K-S} (1 + (1-\rho)(K-S))}{(1-\rho)^2} \right) \\
 &= \frac{P_0 S^S \rho^{S+1}}{S! (1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))].
 \end{aligned}$$

Quando $\rho = 1$,

$$\begin{aligned}
 L_q &= \sum_{n=S+1}^K (n-S) \frac{S^S}{S!} P_0 = \frac{S^S}{S!} P_0 \sum_{n=S+1}^K (n-S) \\
 &= \frac{S^S}{S!} P_0 \sum_{j=1}^{K-S} j = \frac{S^S}{S!} P_0 \frac{(K-S)(K-S+1)}{2}.
 \end{aligned}$$

Concluimos que

$$L_q = \begin{cases} \frac{P_0 S^S \rho^{S+1}}{S! (1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))], & \text{se } \rho \neq 1 \\ \frac{S^S}{S!} P_0 \frac{(K-S)(K-S+1)}{2}, & \text{se } \rho = 1 \end{cases},$$

onde P_0 tem a expressão dada por (24).

Tendo o número médio de clientes, podemos calcular o número médio de clientes no sistema e o tempo médio esperado na fila de cada cliente.

Vamos então, em primeiro lugar, calcular o número médio de clientes no sistema.

Para $\rho \neq 1$,

$$\begin{aligned}
 L_s &= L_q + \frac{\lambda'}{\mu} = \frac{P_0 S^S \rho^{S+1}}{S! (1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))] + \frac{\lambda \left(1 - \frac{S^S \rho^K}{S!} P_0\right)}{\mu} \\
 &= \frac{P_0 S^S \rho^{S+1}}{S! (1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))] + S\rho \left(1 - \frac{S^S \rho^K}{S!} P_0\right) \\
 &= \frac{P_0 S^S \rho^{S+1}}{S! (1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))] + S\rho - \frac{S^{S+1} \rho^{K+1}}{S!} P_0 \\
 &= \frac{P_0 S^S \rho^{S+1}}{S! (1-\rho)^2} \left[1 - \rho^{K-S} (1 + (1-\rho)(K-S)) - S\rho^{K-S} (1-\rho)^2\right] + S\rho \\
 &= \frac{P_0 S^S \rho^{S+1}}{S! (1-\rho)^2} \left[1 - \rho^{K-S} \left(1 + (1-\rho)(K-S) + S(1-\rho)^2\right)\right] + S\rho \\
 &= \frac{P_0 S^S \rho^{S+1}}{S! (1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K-S + S(1-\rho)))] + S\rho \\
 &= \frac{P_0 S^S \rho^{S+1}}{S! (1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K - \rho S))] + S\rho.
 \end{aligned}$$

Quando $\rho = 1$,

$$\begin{aligned}
 L_s &= L_q + \frac{\lambda'}{\mu} = \frac{S^S}{S!} P_0 \frac{(K-S)(K-S+1)}{2} + S \left(1 - \frac{S^S}{S!} P_0\right) \\
 &= \frac{S^S}{S!} P_0 \frac{(K-S)(K-S+1)}{2} + S - \frac{S^{S+1}}{S!} P_0 \\
 &= \frac{S^S}{S!} P_0 \frac{(K-S)(K-S+1) - S}{2} + S.
 \end{aligned}$$

Resumindo,

$$L_s = \begin{cases} \frac{P_0 S^S \rho^{S+1}}{S! (1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K - \rho S))] + S\rho, & \text{se } \rho \neq 1 \\ \frac{S^S}{S!} P_0 \frac{(K-S)(K-S+1) - S}{2} + S, & \text{se } \rho = 1 \end{cases}.$$

Representamos L_s e L_q , como função de ρ , supondo que $K = 30$ e $S = 10$, na Figura 28.

Na Figura 28, podemos observar que L_q é “análoga” a L_s , com umas pequenas adaptações, tendo L_q como limite $K - S$ e tendo L_s como limite K . De facto,

$$\begin{aligned}
 \lim_{\rho \rightarrow +\infty} L_q &= \lim_{\rho \rightarrow +\infty} \left[\frac{S^S \rho^{S+1} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))]}{\left[\sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \frac{(S\rho)^S}{S!} \left(\frac{1 - \rho^{K-S+1}}{1-\rho} \right) \right] S! (1-\rho)^2} \right] \\
 &= \lim_{\rho \rightarrow +\infty} \left[\frac{K S^S \rho^{K+2} - S^{S+1} \rho^{K+2}}{-\frac{S^S \rho^{K+3}}{1-\rho}} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{(K \rho^{K+2} - S \rho^{K+2}) (1-\rho)}{-\rho^{K+3}} \right] \\
 &= \lim_{\rho \rightarrow +\infty} \left[\frac{(K-S)(1-\rho)}{-\rho} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{(K-S) \left(\frac{1}{\rho} - 1 \right)}{-1} \right] \\
 &= K - S.
 \end{aligned}$$

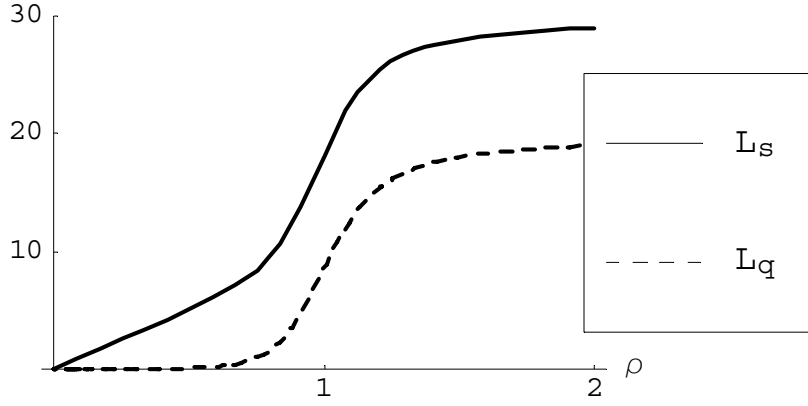


Figura 28: Número médio de clientes no sistema (L_s) e na fila (L_q) em função da taxa de pressão (ρ), com $S = 10$ e $H = 30$

e

$$\begin{aligned}
 \lim_{\rho \rightarrow +\infty} L_s &= \lim_{\rho \rightarrow +\infty} L_q + \lim_{\rho \rightarrow +\infty} \frac{\lambda'}{\mu} = K - S + \lim_{\rho \rightarrow +\infty} S\rho \left(1 - \frac{S^S \rho^K}{S!} P_0 \right) \\
 &= K - S + \lim_{\rho \rightarrow +\infty} S\rho \left[1 - \frac{S^S \rho^K}{S! \left(\sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \frac{(S\rho)^S}{S!} \left(\frac{1 - \rho^{K-S+1}}{1 - \rho} \right) \right)} \right] \\
 &= K - S + \lim_{\rho \rightarrow +\infty} S\rho \left[1 - \frac{S^S \rho^K}{S! \left(\frac{S^S \rho^{K+1}}{S!(1-\rho)} \right)} \right] \\
 &= K - S + \lim_{\rho \rightarrow +\infty} \left[S\rho - S\rho^{K+1} \left(\frac{1-\rho}{\rho^{K+1}} \right) \right] \\
 &= K - S + \lim_{\rho \rightarrow +\infty} [S\rho - S - S\rho] = K - S + S \\
 &= K.
 \end{aligned}$$

Determinamos agora o tempo médio de espera pelo cliente numa fila

$$\begin{aligned}
 W_q &= \frac{L_q}{\lambda'} = \frac{\frac{P_0 S^S \rho^{S+1}}{S!(1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))]}{\lambda \left(1 - \frac{S^S \rho^K}{S!} P_0 \right)} \\
 &= \frac{\frac{P_0 S^{S-1} \rho^S}{S!(1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))]}{\mu \left(1 - \frac{S^S \rho^K}{S!} P_0 \right)} \\
 &= \frac{P_0 S^{S-1} \rho^S}{\mu (1-\rho)^2 (S! - S^S \rho^K P_0)} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))],
 \end{aligned}$$

se $\rho \neq 1$.

Quando $\rho = 1$,

$$\begin{aligned} W_q &= \frac{L_q}{\lambda'} = \frac{\frac{S^S}{S!} P_0 \frac{(K-S)(K-S+1)}{2}}{\lambda \left(1 - \frac{S^S}{S!} P_0\right)} = \frac{\frac{S^S}{S!} P_0 \frac{(K-S)(K-S+1)}{2}}{S\mu \left(1 - \frac{S^S}{S!} P_0\right)} \\ &= \frac{S^{S-1} P_0 (K-S)(K-S+1)}{2\mu (S! - S P_0)}. \end{aligned}$$

Então,

$$W_q = \begin{cases} \frac{P_0 S^{S-1} \rho^S}{\mu (1-\rho)^2 (S! - S^S \rho^K P_0)} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))], & \text{se } \rho \neq 1 \\ \frac{S^{S-1} P_0 (K-S)(K-S+1)}{2\mu (S! - S P_0)}, & \text{se } \rho = 1 \end{cases}.$$

Supondo que $S = 10$ e que $K = 30$, representamos W_q como função de ρ na Figura 29, para os casos $\mu = 0.5, 1, 5$ e 50 .

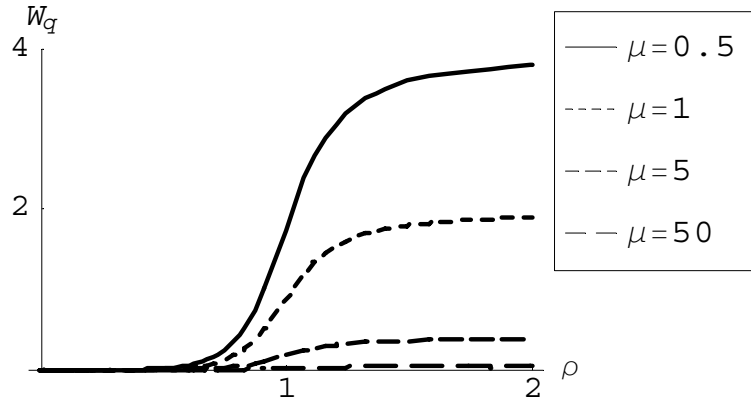


Figura 29: Tempo médio de espera na fila (W_q) em função da taxa de pressão (ρ), com $S = 10$ e $H = 30$

Podemos observar na Figura 29 que, independentemente da taxa de serviço, o tempo médio de espera na fila aumenta consoante aumenta a taxa de pressão. Verifica-se que consoante aumenta a taxa de serviço, o tempo de espera diminui drasticamente. Podemos verificar que tal como se gere o gráfico, W_q está limitado, quando K , S e μ são fixos. De facto, tendo em conta (24) e que $S < K$ vem

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} W_q &= \lim_{\rho \rightarrow +\infty} \left[\frac{P_0 S^{S-1} \rho^S [1 - \rho^{K-S} (1 + (1-\rho)(K-S))]}{\mu (1-\rho)^2 (S! - S^S \rho^K P_0)} \right] \\ &= \lim_{\rho \rightarrow +\infty} \left[\frac{S^{S-1} \rho^S [1 - \rho^{K-S} (1 + (1-\rho)(K-S))]}{\left(-\frac{S^S \rho^{K+1}}{S!(1-\rho)}\right) \mu (1-\rho)^2 \left(S! - \frac{S^S \rho^K}{-\frac{S^S \rho^{K+1}}{S!(1-\rho)}}\right)} \right] \\ &= \lim_{\rho \rightarrow +\infty} \left[\frac{K S^{S-1} \rho^{K+1} - S^S \rho^{K+1}}{\left(-\frac{S^S \rho^{K+1}}{S!}\right) \mu (1-\rho) \left(S! + \frac{S^S \rho^K S! (1-\rho)}{S^S \rho^{K+1}}\right)} \right] \\ &= \lim_{\rho \rightarrow +\infty} \left[\frac{S^S \rho^{K+1} (K-1)}{-S^S S \rho^K \mu (1-\rho)} \right] = \frac{K-1}{S\mu}. \end{aligned}$$

Sabemos que o tempo esperado no sistema pode ser determinado de diferentes maneiras, nomeadamente,

$$W_s = W_q + \frac{1}{\mu} = \frac{L_s}{\lambda'}.$$

Vamos calculá-lo através de W_q . Quando $\rho \neq 1$,

$$\begin{aligned} W_s &= W_q + \frac{1}{\mu} \\ &= \frac{P_0 S^{S-1} \rho^S}{\mu (1-\rho)^2 (S! - S^S \rho^K P_0)} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))] + \frac{1}{\mu}. \end{aligned}$$

Quando $\rho = 1$,

$$\begin{aligned} W_s &= W_q + \frac{1}{\mu} \\ &= \frac{S^{S-1} P_0 (K-S)(K-S+1)}{2\mu (S! - S P_0)} + \frac{1}{\mu}. \end{aligned}$$

Ou seja,

$$W_s = \begin{cases} \frac{P_0 S^{S-1} \rho^S}{\mu (1-\rho)^2 (S! - S^S \rho^K P_0)} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))] + \frac{1}{\mu}, & \text{se } \rho \neq 1 \\ \frac{S^{S-1} P_0 (K-S)(K-S+1)}{2\mu (S! - S P_0)} + \frac{1}{\mu}, & \text{se } \rho = 1 \end{cases}$$

Supondo novamente que $K = 30$ e $S = 10$, representamos W_s como função de ρ na Figura 30, para os casos $\mu = 0.5, 1, 5$ e 50 .

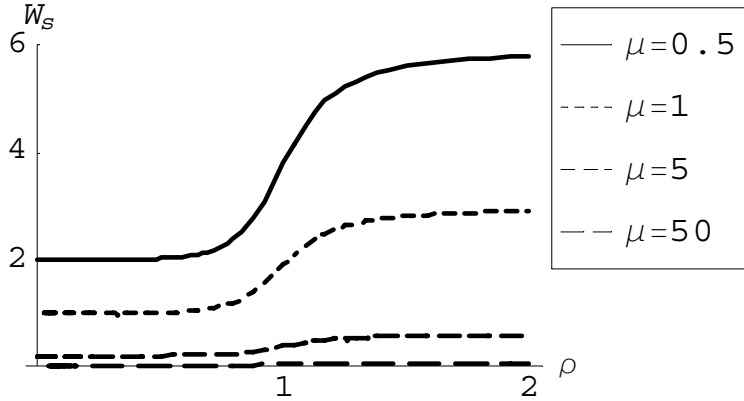


Figura 30: Tempo médio de espera no sistema (W_s) em função da taxa de pressão (ρ), com $S = 10$ e $H = 30$

A Figura 30 difere pouco da Figura 29. A grande diferença tem a ver com o facto de os gráficos não começarem em 0, pois o tempo de serviço do cliente que está a ser servido é contabilizado em W_s . De facto,

$$\begin{aligned} \lim_{\rho \rightarrow 0} W_s &= \lim_{\rho \rightarrow 0} \left[\frac{P_0 S^{S-1} \rho^S}{\mu (1-\rho)^2 (S! - S^S \rho^K P_0)} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))] + \frac{1}{\mu} \right] \\ &= \frac{1}{\mu}. \end{aligned}$$

Quando $\rho \rightarrow +\infty$ os gráficos sugerem a existência de um limite de W_s . De facto,

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} W_s &= \lim_{\rho \rightarrow +\infty} \left[W_q + \frac{1}{\mu} \right] = \frac{K-1}{S\mu} + \frac{1}{\mu} \\ &= \frac{K+S-1}{S\mu}. \end{aligned}$$

A probabilidade de o tempo de espera na fila ser zero, $P[T_q = 0]$, é a probabilidade de o sistema ter $S-1$ ou menos clientes

$$P[T_q = 0] = \sum_{n=0}^{S-1} P_n = \sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} P_0.$$

Esta probabilidade está representada, para $K = 30$ e $S = 10$, na Figura 31.

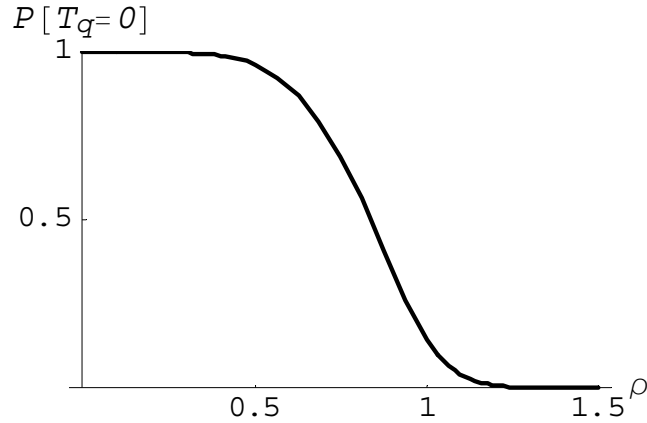


Figura 31: Probabilidade de o tempo de espera na fila ser zero, $P[T_q = 0]$, em função da taxa de pressão, ρ , com $S = 10$ e $K = 30$

Analisando a Figura 31, podemos verificar que a probabilidade de um cliente chegar ao sistema e não ter que esperar no sistema é 1 quando a taxa de pressão é zero e mantém-se perto de 1 até sensivelmente quando $\rho = \frac{1}{2}$ e depois converge para 0 quando a taxa de pressão tende para infinito. Este último resultado é válido, qualquer que sejam os valores de S e de K pois

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} P[T_q = 0] &= \lim_{\rho \rightarrow +\infty} \left[\sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} P_0 \right] \\ &= \lim_{\rho \rightarrow +\infty} \left[\frac{(S\rho)^{S-1}}{(S-1)! \left(\sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \frac{(S\rho)^S}{S!} \left(\frac{1-\rho^{K-S+1}}{1-\rho} \right) \right)} \right] \\ &= \lim_{\rho \rightarrow +\infty} \left[\frac{S^{S-1} \rho^{S-1}}{(S-1)! \frac{S^S \rho^S}{S!} \left(\frac{1-\rho^{K-S+1}}{1-\rho} \right)} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{1}{\frac{S\rho}{S} \left(\frac{1-\rho^{K-S+1}}{1-\rho} \right)} \right] \\ &= \lim_{\rho \rightarrow +\infty} \left[\frac{1-\rho}{\rho - \rho^{K-S+2}} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{\frac{1}{\rho^{K-S+2}} - \frac{\rho}{\rho^{K-S+2}}}{\rho^{K-S+2} - 1} \right] \\ &= 0. \end{aligned}$$

O sistema pode atingir o equilíbrio mesmo para valores de ρ superiores a 1. No entanto, nesse caso, haverá um grande número de clientes que chegam e não são atendidos.

O quadro que se segue é o quadro resumo do modelo $M/M/S/K$.

Probabilidade de ocorrência do estado 0	$P_0 = \begin{cases} \left[\sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} + \frac{(S\rho)^S}{S!} \left(\frac{1-\rho^{K-S+1}}{1-\rho} \right) \right]^{-1}, & \rho \neq 1 \\ \left[\sum_{n=0}^{S-1} \frac{S^n}{n!} + \frac{S^S}{S!} (K-S+1) \right]^{-1}, & \rho = 1 \end{cases}$
Probabilidade de ocorrência do estado n	$P_n = \begin{cases} \frac{(S\rho)^n}{n!} P_0, & n = 0, 1, \dots, S-1 \\ \frac{S^S \rho^n}{S!} P_0, & n = S, \dots, K \\ 0, & n > K \end{cases}$
Número médio de clientes no sistema	$L_s = \begin{cases} S\rho + \frac{P_0 S^S \rho^{S+1}}{S!(1-\rho)^2} \times \\ \times [1 - \rho^{K-S} (1 + (1-\rho)(K-\rho S))] , & \rho \neq 1 \\ \frac{S^S}{S!} P_0 \frac{(K-S)(K-S+1)-S}{2} + S, & \rho = 1 \end{cases}$
Comprimento médio da fila	$L_q = \begin{cases} \frac{P_0 S^S \rho^{S+1}}{S!(1-\rho)^2} [1 - \rho^{K-S} (1 + (1-\rho)(K-S))] , & \rho \neq 1 \\ \frac{S^S}{S!} P_0 \frac{(K-S)(K-S+1)}{2}, & \rho = 1 \end{cases}$
Tempo médio de espera no sistema	$W_s = \begin{cases} \frac{1}{\mu} + \frac{P_0 S^{S-1} \rho^S}{\mu(1-\rho)^2 (S! - S^S \rho^K P_0)} \times \\ \times [1 - \rho^{K-S} (1 + (1-\rho)(K-S))] , & \rho \neq 1 \\ \frac{S^{S-1} P_0 (K-S)(K-S+1)}{2\mu(S! - S P_0)} + \frac{1}{\mu}, & \rho = 1 \end{cases}$
Tempo médio de espera na fila	$W_q = \begin{cases} \frac{P_0 S^{S-1} \rho^S}{\mu(1-\rho)^2 (S! - S^S \rho^K P_0)} \times \\ \times [1 - \rho^{K-S} (1 + (1-\rho)(K-S))] , & \rho \neq 1 \\ \frac{S^{S-1} P_0 (K-S)(K-S+1)}{2\mu(S! - S P_0)}, & \rho = 1 \end{cases}$
Prob. do tempo de espera na fila ser 0	$P(T_q = 0) = \sum_{n=0}^{S-1} \frac{(S\rho)^n}{n!} P_0$
Taxa de pressão	$\rho = \frac{\lambda}{S\mu}$

Quadro 5: Sumário das características do modelo $M/M/S/K$

4.5.1 Modelo $M/M/S/S$

Este modelo é, digamos, o caso limite do modelo $M/M/S/K$, onde se supõe que $S < K$. É um dos primeiros modelos estudados por Erlang, para analisar a probabilidade de n linhas estarem ocupadas numa central telefônica que dispõe de S linhas (cf. [1]). Neste caso, a capacidade $K = S$, e se uma chamada chega à central quando todas as linhas estão ocupadas, o cliente não permanece à espera pois recebe um sinal de saturação e a comunicação é abortada nesse momento. Por outras palavras, o número de posições em serviço é igual ao número de servidores (não há fila de espera).

Neste caso a taxa de chegada é

$$\lambda_n = \begin{cases} \lambda & \text{se } n = 0, 1, 2, \dots, S-1 \\ 0 & \text{se } n \geq S \end{cases}$$

e a taxa de atendimento é

$$\mu_n = \begin{cases} n\mu & \text{se } n = 0, 1, 2, \dots, S \\ 0 & \text{se } n \geq S+1 \end{cases}.$$

E do modelo anterior temos que,

$$\lambda' = \lambda(1 - P_S).$$

Note-se que este sistema poderá ser representado através da Figura 32.

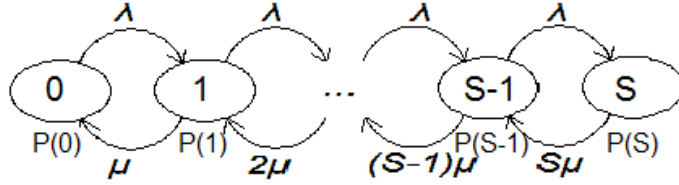


Figura 32: Modelo $M/M/S/S$

A probabilidade P_S de que o sistema se encontre cheio e que ao chegar uma ligação esta se perca é dada pela chamada fórmula de perda da distribuição de Erlang (cf. [6]), que se obtém tomando em (24) e em (23) $K = S$:

$$P_n = \frac{\frac{(\lambda/\mu)^n}{n!}}{\sum_{j=0}^S \frac{(\lambda/\mu)^j}{j!}} = \frac{\frac{(S\rho)^n}{n!}}{\sum_{j=0}^S \frac{(S\rho)^j}{j!}},$$

para $n = 0, 1, 2, \dots, S$.

Representamos a fórmula de perda da distribuição de Erlang graficamente, na Figura 33 (com ρ a variar de 0 a 1.5) e na Figura 34 (com ρ a variar de 0 a 10), supondo, em ambos os casos, que $S = 20$.

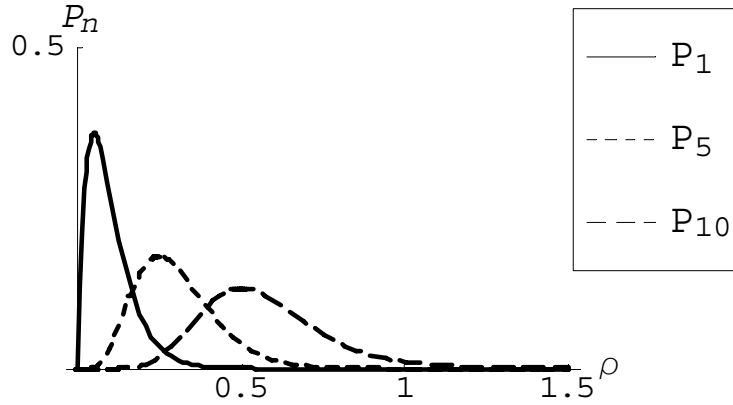


Figura 33: Probabilidade de existirem n clientes no sistema (P_n) em função da taxa de pressão (ρ) quando este varia de 0 a 1.5, com $S = 20$

Da Figura 33, podemos verificar que independentemente do número de clientes no sistema (desde que não nulo), quando ρ toma o valor limite a probabilidade é 0, aumenta um pouco

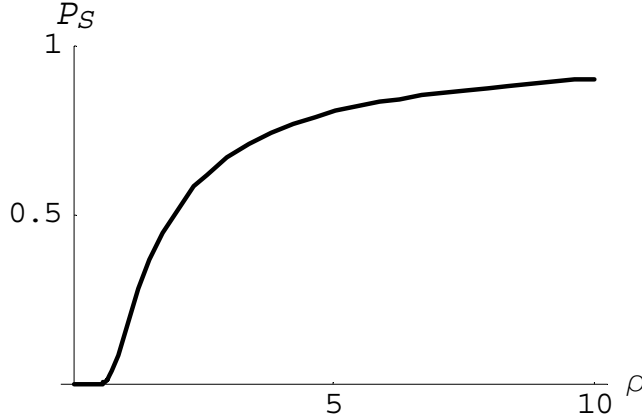


Figura 34: Probabilidade de existirem S clientes no sistema (P_S) em função da taxa de pressão (ρ) quando este varia de 0 a 10, com $S = 20$

e depois converge para zero. Com a Figura 34 verificamos que quando ρ tende para infinito a probabilidade de existir S clientes tende para 1. De facto,

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} P_S &= \lim_{\rho \rightarrow +\infty} \left[\frac{\frac{(\lambda/\mu)^S}{S!}}{\sum_{j=0}^S \frac{(\lambda/\mu)^j}{j!}} \right] = \lim_{\rho \rightarrow +\infty} \left[\frac{\frac{S^S \rho^S}{S!}}{\frac{S^S \rho^S}{S!}} \right] \\ &= 1. \end{aligned}$$

Como já foi referido, não existe espaço para os clientes estarem numa "fila de espera". Por isso, obviamente, o tamanho médio da fila de espera, L_q , e o tempo médio de espera de um cliente na fila, W_q , é zero, ou seja

$$L_q = W_q = 0.$$

Usando a formula de Little temos que

$$\begin{aligned} W_s &= W_q + \frac{1}{\mu} \\ &= \frac{1}{\mu}, \end{aligned}$$

tal como acontecia no modelo $M/M/\infty$. A Figura 35 é, por isso, igual à Figura 17.

Usando novamente as fórmulas de Little temos

$$\begin{aligned} L_s &= L_q + \frac{\lambda'}{\mu} = \frac{\lambda(1 - P_S)}{\mu} \\ &= S\rho(1 - P_S). \end{aligned}$$

Na Figura 36 apresentamos o gráfico do número médio de clientes no sistema, supondo $S = 20$.

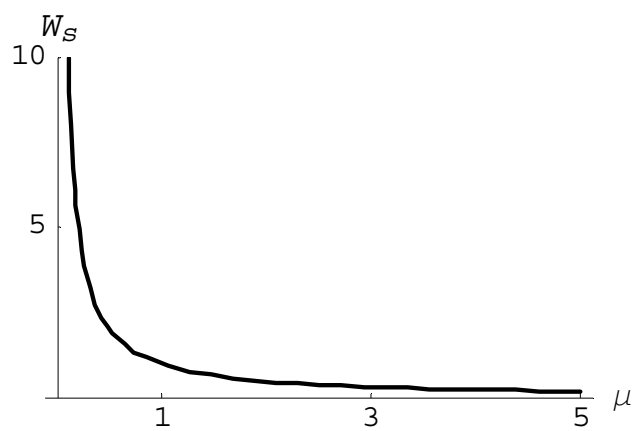


Figura 35: Tempo médio de espera no sistema (W_s) em função da taxa de serviço (μ)

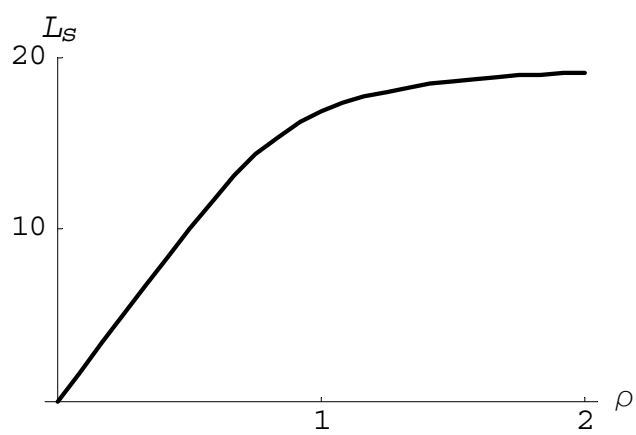


Figura 36: Número médio de clientes no sistema (L_s) em função da taxa de pressão (ρ), com $S = 20$

Analisando a Figura 36, verificamos que quando a taxa de pressão é zero, o número médio de clientes também é zero, e quando a taxa de pressão converge para infinito, o número médio de clientes tende para S

$$\begin{aligned}
 \lim_{\rho \rightarrow +\infty} L_s &= \lim_{\rho \rightarrow +\infty} S\rho(1 - P_S) = \lim_{\rho \rightarrow +\infty} S\rho \left(1 - \frac{\frac{(S\rho)^S}{S!}}{\sum_{j=0}^S \frac{(S\rho)^j}{j!}} \right) \\
 &= \lim_{\rho \rightarrow +\infty} S\rho \left(\frac{\sum_{j=0}^{S-1} \frac{(S\rho)^j}{j!}}{\sum_{j=0}^S \frac{(S\rho)^j}{j!}} \right) = \lim_{\rho \rightarrow +\infty} S\rho \frac{(S\rho)^{S-1} S!}{(S-1)! (S\rho)^S} = \lim_{\rho \rightarrow +\infty} \frac{(S\rho)^S S}{(S\rho)^S} \\
 &= S.
 \end{aligned}$$

O quadro que se segue é o quadro reúne as características deste modelo.

Probabilidade de ocorrência do estado 0	$P_0 = \left[\sum_{j=0}^S \frac{(S\rho)^j}{j!} \right]^{-1}$
Probabilidade de ocorrência do estado n	$P_n = \begin{cases} \frac{(S\rho)^n}{n!}, & n = 0, \dots, S \\ 0, & n > S \end{cases}$
Número médio de clientes no sistema	$L_s = S\rho(1 - P_S)$
Comprimento médio da fila	$L_q = 0$
Tempo médio de espera no sistema	$W_s = \frac{1}{\mu}$
Tempo médio de espera na fila	$W_q = 0$
Taxa de pressão	$\rho = \frac{\lambda}{S\mu}$

Quadro 6: Sumário das características do modelo $M/M/S/S$

4.6 Modelo $M/M/1/\infty/H$

Neste modelo, o sistema tem uma distribuição dos tempos entre as chegadas e dos tempos de atendimento exponencial, contém 1 servidor, a capacidade do sistema é infinita, a capacidade da população é de H indivíduos e a disciplina corresponde a quem entra primeiro no sistema ser o primeiro a ser atendido e a sair. Este modelo pode também ser representado por $M/M/1/\infty/H/FIFO$ ou por $M/M/1/\infty/H/FCFS$. Note-se que o comportamento deste modelo não é alterado se substituirmos a capacidade infinita por uma capacidade finita, desde que esta não seja inferior à dimensão da população, ou seja, este modelo pode ser transformado no modelo $M/M/1/K/H$ onde $K \in \{H, H+1, \dots\} \cup \{+\infty\}$.

A distribuição do tempo entre as chegadas de clientes ao sistema dependerá de quantos clientes estão no sistema. Considera-se que, para cada cliente que acaba de sair do sistema, a distribuição de probabilidade do tempo que falta para que volte a entrar é $\exp(\lambda)$.

Se o número de clientes no sistema for $N = n$, o número de clientes na população será $H - n$. A taxa de chegadas é

$$\lambda_n = \begin{cases} \lambda(H - n), & \text{se } n = 0, 1, 2, \dots, H \\ 0, & \text{se } n > H \end{cases},$$

e a taxa de de serviço é

$$\mu_n = \mu, \quad \text{se } n > 1.$$

Podemos representar este sistema através da Figura 37.

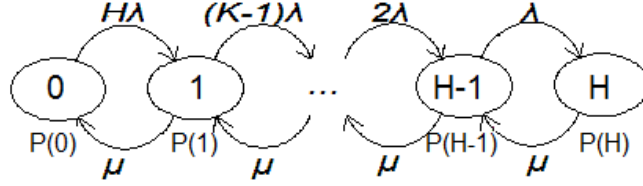


Figura 37: Modelo $M/M/1/\infty/H$

Resulta, tomando $\rho = \frac{\lambda}{\mu}$, pois

$$\begin{aligned} \theta_n &= \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} = \frac{\lambda H \lambda (H-1) \dots \lambda (H-n+1)}{\mu^n} \\ &= \frac{\lambda^n H (H-1) \dots (H-n+1)}{\mu^n} \\ &= \rho^n H (H-1) \dots (H-n+1), \end{aligned}$$

$$P_n = \begin{cases} \frac{H!}{(H-n)!} \rho^n P_0, & \text{se } n = 0, 1, 2, \dots, H \\ 0, & \text{se } n > H \end{cases}, \quad (25)$$

onde

$$P_0 = \left[\sum_{n=0}^H \frac{H!}{(H-n)!} \rho^n \right]^{-1}. \quad (26)$$

Supondo que $H = 30$, representamos graficamente P_n na Figura 38 quando ρ varia entre 0 e 0.5, e na Figura 39 quando ρ varia entre 0 e 10.

As Figuras 38 e 39 são semelhantes às Figuras 26 e 27. Mais uma vez, verificamos que quando se trata da probabilidade de existirem H clientes no sistema, quando ρ tende para o infinito, a probabilidade converge para 1. De facto,

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} P_H &= \lim_{\rho \rightarrow +\infty} \left(\frac{H! \rho^H}{\sum_{n=0}^H \frac{H!}{(H-n)!} \rho^n} \right) = \lim_{\rho \rightarrow +\infty} \left(\frac{H! \rho^H}{H! \rho^H} \right) \\ &= 1. \end{aligned}$$

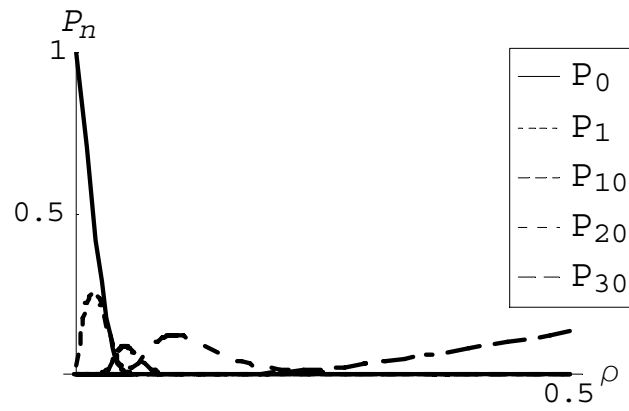


Figura 38: Probabilidade de existirem n clientes no sistema (P_n) em função da taxa de pressão (ρ) quando esta varia de 0 a 0.5, com $H = 30$

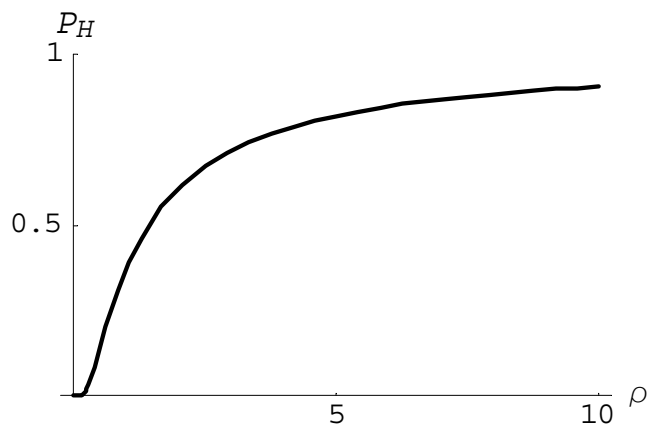


Figura 39: Probabilidade de existirem H clientes no sistema (P_n) em função da taxa de pressão (ρ) quando esta varia de 0 a 10, com $H = 30$

Relativamente ao número médio de clientes no sistema,

$$L_s = \sum_{n=0}^H n P_n,$$

temos

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} L_s &= \lim_{\rho \rightarrow +\infty} \sum_{n=0}^H n P_n = \lim_{\rho \rightarrow +\infty} \sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n P_0 = \lim_{\rho \rightarrow +\infty} P_0 \sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n \\ &= \lim_{\rho \rightarrow +\infty} \frac{1}{\sum_{n=0}^H \frac{H!}{(H-n)!} \rho^n} \sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n = \lim_{\rho \rightarrow +\infty} \frac{1}{\frac{H!}{(H-H)!} \rho^H} H \frac{H!}{(H-H)!} \rho^H \\ &= H. \end{aligned} \quad (27)$$

Temos ainda que

$$\begin{aligned} \lambda' &= \sum_{n=0}^H \lambda_n P_n = \sum_{n=0}^H \lambda (H-n) P_n = \lambda \left(\sum_{n=0}^H H P_n - \sum_{n=0}^H n P_n \right) \\ &= \lambda (H - L_s). \end{aligned} \quad (28)$$

Tendo isto em conta resulta que

$$\begin{aligned} W_s &= \frac{L_s}{\lambda'} = \frac{L_s}{\lambda (H - L_s)} \\ &= \frac{L_s}{\rho \mu (H - L_s)}. \end{aligned} \quad (29)$$

Supondo que $H = 30$, o gráfico da Figura 40 corresponde ao tempo médio de espera no sistema como função de ρ , para $\mu = 0.5, 1, 5$ e 50 .

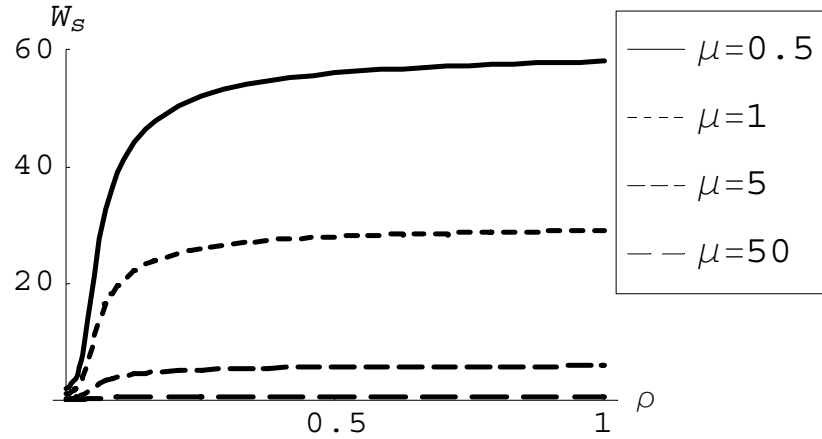


Figura 40: Tempo médio de espera no sistema (W_s) em função da taxa de pressão (ρ), com $H = 30$

Observando a Figura 40, verificamos que independentemente da taxa de serviço, o tempo de espera, quando a taxa de pressão tende para o infinito, estabiliza, tendo como limite de

$\frac{H}{\mu}$:

$$\begin{aligned}
\lim_{\rho \rightarrow +\infty} W_s &= \lim_{\rho \rightarrow +\infty} \frac{L_s}{\rho\mu(H-L_s)} = \lim_{\rho \rightarrow +\infty} \frac{\sum_{n=0}^H nP_n}{\rho\mu\left(H - \sum_{n=0}^H nP_n\right)} \\
&= \lim_{\rho \rightarrow +\infty} \frac{\sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n P_0}{\rho\mu\left(H - \sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n P_0\right)} = \lim_{\rho \rightarrow +\infty} \frac{\sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n}{\rho\mu\left(\frac{H}{P_0} - \sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n\right)} \\
&= \lim_{\rho \rightarrow +\infty} \frac{\sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n}{\rho\mu\left(\sum_{n=0}^H H \frac{H!}{(H-n)!} \rho^n - \sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n\right)} \\
&= \lim_{\rho \rightarrow +\infty} \frac{\sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n}{\rho\mu \sum_{n=0}^{H-1} \frac{H!}{(H-n)!} \rho^n (H-n)} = \lim_{\rho \rightarrow +\infty} \frac{H \frac{H!}{(H-H)!} \rho^H}{\rho\mu \frac{H!}{(H-(H-1))!} \rho^{H-1} (H-(H-1))} \\
&= \frac{H}{\mu}.
\end{aligned}$$

Temos que o número médio que um cliente espera na fila é,

$$\begin{aligned}
W_q &= W_s - \frac{1}{\mu} = \frac{L_s}{\lambda(H-L_s)} - \frac{1}{\mu} \\
&= \frac{L_s}{\rho\mu(H-L_s)} - \frac{1}{\mu}.
\end{aligned} \tag{30}$$

Que podemos representar na Figura 41, supondo que $H = 30$.

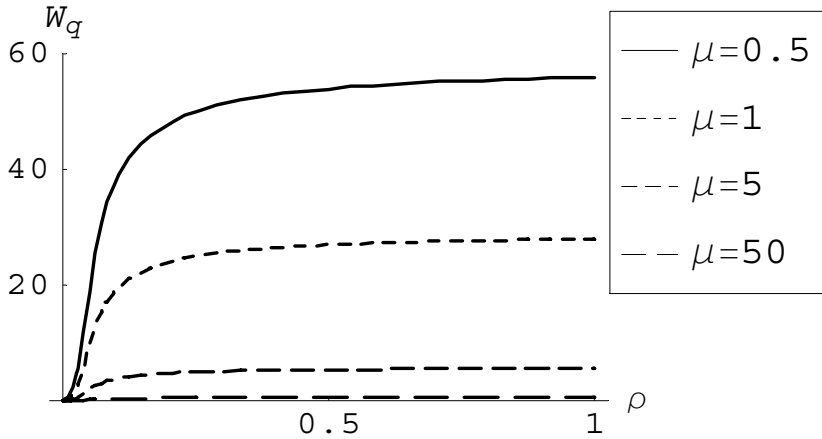


Figura 41: Tempo médio de espera no sistema (W_q) em função da taxa de pressão (ρ), com $H = 30$

Como era de esperar, os gráficos da Figura 41 são praticamente idênticos aos da Figura 40 e convergem para $\frac{H-1}{\mu}$,

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} W_q &= \lim_{\rho \rightarrow +\infty} \left(W_s - \frac{1}{\mu} \right) = \frac{H}{\mu} - \frac{1}{\mu} \\ &= \frac{H-1}{\mu}. \end{aligned}$$

O número médio de clientes na fila é (cf. (28) e (30))

$$\begin{aligned} L_q &= \lambda' W_q = \lambda (H - L_s) \left(\frac{L_s}{\lambda (H - L_s)} - \frac{1}{\mu} \right) \\ &= L_s - \frac{\lambda (H - L_s)}{\mu} \\ &= L_s - \rho (H - L_s). \end{aligned} \tag{31}$$

Representamos graficamente na Figura 42, supondo que $H = 30$, L_s e L_q .

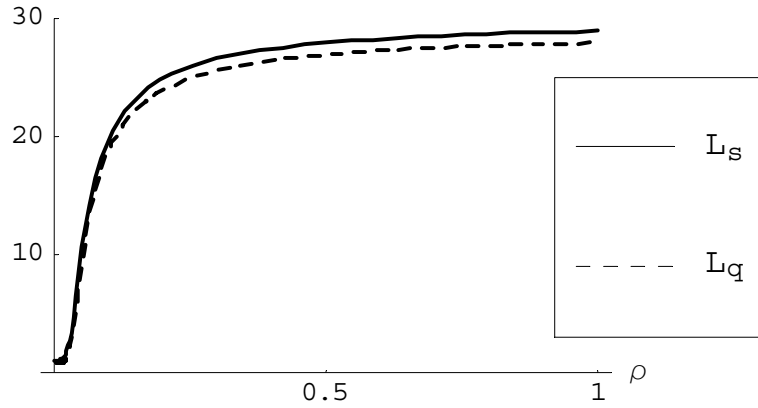


Figura 42: Número médio de clientes no sistema (L_s) e na fila (L_q) em função da taxa de pressão (ρ), com $H = 30$

O número médio de clientes no sistema e na fila é pequeno quando a taxa de pressão é zero, mas depois aumenta, sem nunca atingir H , como mostra a Figura 42. De facto, de (27), sabemos que L_s converge para H . Usando esse facto e (31) temos

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} L_q &= \lim_{\rho \rightarrow +\infty} [L_s - \rho (H - L_s)] \\ &= \lim_{\rho \rightarrow +\infty} L_s - \lim_{\rho \rightarrow +\infty} \rho (H - L_s) \end{aligned}$$

de (25) e (26),

$$\begin{aligned}
\lim_{\rho \rightarrow +\infty} L_q &= H - \lim_{\rho \rightarrow +\infty} \rho (H - L_s) = H - \lim_{\rho \rightarrow +\infty} \rho \left(H - \sum_{n=0}^H n P_n \right) \\
&= H - \lim_{\rho \rightarrow +\infty} \rho \left[H - \sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n P_0 \right] \\
&= H - \lim_{\rho \rightarrow +\infty} \rho \left[H - \sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n \frac{1}{\sum_{n=0}^H \frac{H!}{(H-n)!} \rho^n} \right] \\
&= H - \lim_{\rho \rightarrow +\infty} \rho \left[\frac{\sum_{n=0}^H H \frac{H!}{(H-n)!} \rho^n - \sum_{n=0}^H n \frac{H!}{(H-n)!} \rho^n}{\sum_{n=0}^H \frac{H!}{(H-n)!} \rho^n} \right] \\
&= H - \lim_{\rho \rightarrow +\infty} \rho \left[\frac{\sum_{n=0}^{H-1} \frac{H!}{(H-n)!} \rho^n (H-n)}{\sum_{n=0}^H \frac{H!}{(H-n)!} \rho^n} \right] \\
&= H - \lim_{\rho \rightarrow +\infty} \rho \frac{\frac{H!}{(H-(H-1))!} \rho^{H-1} (H - (H-1))}{\frac{H!}{(H-H)!} \rho^H} \\
&= H - 1.
\end{aligned}$$

Podemos então resumir num quadro as características do modelo $M/M/1/\infty/H$,

Probabilidade de ocorrência do estado 0	$P_0 = \left[\sum_{n=0}^H \frac{H!}{(H-n)!} \rho^n \right]^{-1}$
Probabilidade de ocorrência do estado n	$P_n = \begin{cases} \frac{H!}{(H-n)!} \rho^n P_0, & n = 0, 1, 2, \dots, H \\ 0, & n > H \end{cases}$
Número médio de clientes no sistema	$L_s = \sum_{n=0}^H n P_n$
Comprimento médio da fila	$L_q = L_s - \rho (H - L_s)$
Tempo médio de espera no sistema	$W_s = \frac{L_s}{\rho \mu (H - L_s)}$
Tempo médio de espera na fila	$W_q = \frac{L_q}{\rho \mu (H - L_s)} - \frac{1}{\mu}$
Taxa de pressão	$\rho = \frac{\lambda}{\mu}$

Quadro 7: Sumário das características do modelo $M/M/1/\infty/H$

4.7 Modelo $M/M/S/\infty/H$

Este modelo corresponde ao modelo em que o sistema tem uma distribuição das chegadas de Poisson e dos tempos de atendimento exponencial, contém S servidores, a capacidade do sistema é infinita, a capacidade da população é de H indivíduos e a disciplina corresponde a quem entra primeiro no sistema ser o primeiro a ser atendido e a sair. Este modelo é uma

generalização do modelo anterior, por isso algumas características anteriores são válidas neste modelo. Tal como no modelo anterior, notemos que o comportamento deste modelo não é alterado se substituirmos a capacidade infinita por uma capacidade finita, desde que esta não seja inferior à dimensão da população, ou seja, este modelo pode ser também apresentado na forma $M/M/S/K/H$ onde $K \in \{H, H+1, \dots\} \cup \{+\infty\}$.

Supomos que $S < H$, pois se $S \geq H$ todos os clientes a chegar ao sistema seriam atendidos de imediato e não haveria fila de espera.

A taxa de chegada é igual ao do modelo anterior,

$$\lambda_n = \begin{cases} \lambda(H-n), & \text{se } n = 0, 1, 2, \dots, H \\ 0, & \text{se } n > H \end{cases}.$$

As taxas de serviço são iguais ao modelo $M/M/S$ ou $M/M/S/K$

$$\mu_n = \begin{cases} n\mu, & \text{se } n = 1, \dots, S-1 \\ S\mu, & \text{se } n = S, \dots, H \end{cases}.$$

Este sistema pode ser representado através da Figura 43.

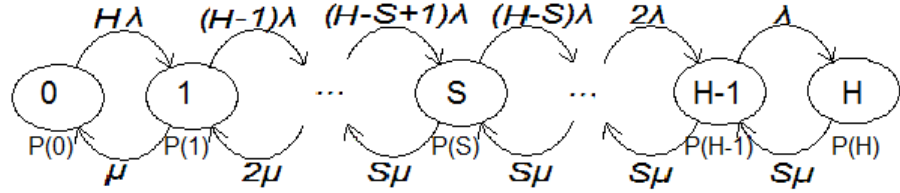


Figura 43: Modelo $M/M/S/\infty/H$

Podemos determinar P_n com a ajuda do modelo anterior e do modelo $M/M/S$

$$P_n = \begin{cases} \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} P_0, & \text{se } n = 1, \dots, S \\ \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n P_0, & \text{se } n = S+1, \dots, H \\ 0, & \text{se } n > H \end{cases},$$

onde

$$P_0 = \left[\sum_{n=0}^S \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n \right]^{-1}.$$

Supondo que $S = 10$ e $H = 30$, representamos P_n graficamente na Figura 44, para $n = 0, 1, 5, 20$ e 30 , quando ρ varia entre 0 e 0.3 e na Figura 45, para $n = H$, quando ρ varia entre 0 e 20.

Os gráficos obtidos são semelhantes aos obtidos na secção anterior. Na Figura 45, tal como no caso anterior, verificamos que a probabilidade de existirem H clientes no sistema quando ρ tende para infinito é 1. De facto,

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} P_H &= \lim_{\rho \rightarrow +\infty} \left[\frac{H! S^S \rho^H}{S! \left(\sum_{n=0}^S \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n \right)} \right] \\ &= \lim_{\rho \rightarrow +\infty} \left[\frac{H! S^S \rho^H}{S! \left(\frac{H! S^S \rho^H}{S!} \right)} \right] \\ &= 1. \end{aligned}$$

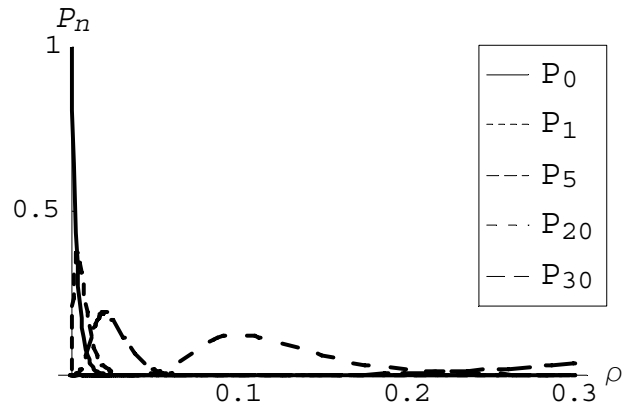


Figura 44: Probabilidade de existirem n clientes no sistema (P_n) em função da taxa de pressão (ρ) que varia entre 0 e 0.3, com $S = 10$ e $H = 30$

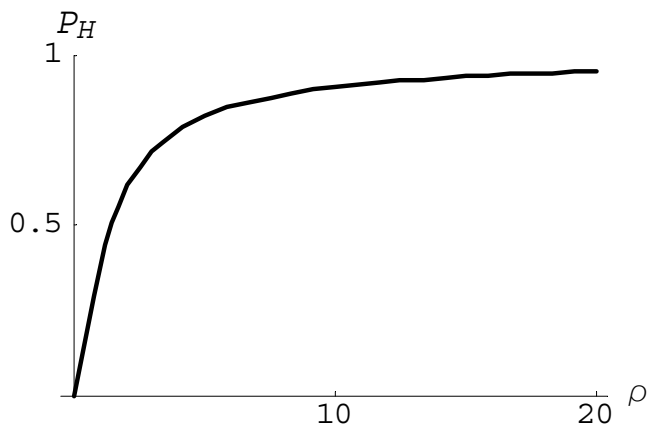


Figura 45: Probabilidade de existirem H clientes no sistema (P_H) em função da taxa de pressão (ρ) que varia entre 0 a 20, com $S = 10$ e $H = 30$

O número médio de clientes no sistema é

$$L_s = \sum_{n=0}^H nP_n.$$

Que é um somatório limitado, pois,

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} L_s &= \lim_{\rho \rightarrow +\infty} \left(\sum_{n=0}^H nP_n \right) \\ &= \lim_{\rho \rightarrow +\infty} P_0 \left[\sum_{n=0}^S n \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H n \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n \right] \\ &= \lim_{\rho \rightarrow +\infty} \left[\frac{\sum_{n=0}^S n \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H n \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n}{\sum_{n=0}^S \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n} \right] \\ &= \lim_{\rho \rightarrow +\infty} \left(\frac{H \frac{H!}{0!} \frac{S^S}{S!} \rho^H}{\frac{H!}{0!} \frac{S^S}{S!} \rho^H} \right) \\ &= H. \end{aligned}$$

Temos ainda que, tal como no caso anterior,

$$\lambda' = \lambda(H - L_s),$$

que

$$\begin{aligned} W_s &= \frac{L_s}{\lambda'} = \frac{L_s}{\lambda(H - L_s)} \\ &= \frac{L_s}{S\rho\mu(H - L_s)}. \end{aligned}$$

Na Figura 46, representamos W_s quando $S = 10$, $H = 30$ e $\mu \in \{0.5, 1, 5, 50\}$.

Analisando os gráficos da Figura 46 concluímos que o tempo de espera aumenta consoante aumenta a taxa de pressão e quanto maior for a taxa de serviço menos tempo o cliente espera

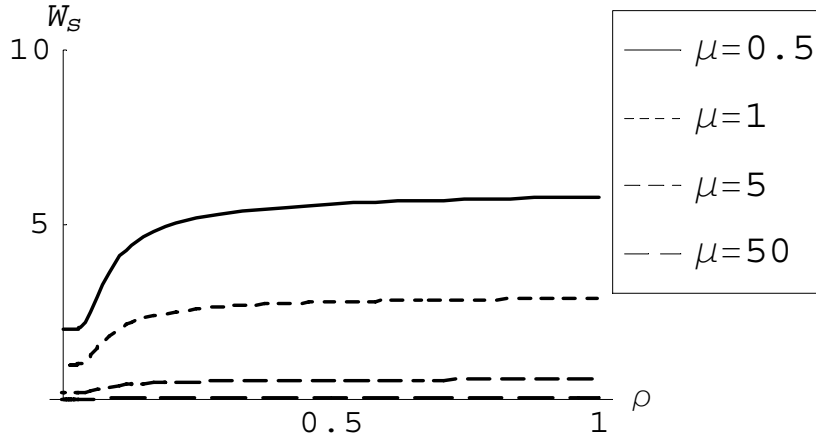


Figura 46: Tempo médio de espera no sistema (W_s) em função da taxa de pressão (ρ), com $S = 10$ e $H = 30$

no sistema. Observamos também que W_s converge. De facto,

$$\begin{aligned}
 \lim_{\rho \rightarrow +\infty} W_s &= \lim_{\rho \rightarrow +\infty} \frac{L_s}{S\rho\mu(H - L_s)} = \lim_{\rho \rightarrow +\infty} \frac{\sum_{n=0}^H nP_n}{S\rho\mu\left(H - \sum_{n=0}^H nP_n\right)} \\
 &= \lim_{\rho \rightarrow +\infty} \frac{\sum_{n=0}^S nP_n + \sum_{n=S+1}^H nP_n}{S\rho\mu\left[H - \left(\sum_{n=0}^S nP_n + \sum_{n=S+1}^H nP_n\right)\right]} \\
 &= \lim_{\rho \rightarrow +\infty} \frac{P_0\left(\sum_{n=0}^S n \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H n \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n\right)}{S\rho\mu\left[H - P_0\left(\sum_{n=0}^S n \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H n \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n\right)\right]} \\
 &= \lim_{\rho \rightarrow +\infty} \frac{\sum_{n=0}^S n \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H n \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n}{S\rho\mu\left[\frac{H}{P_0} - \left(\sum_{n=0}^S n \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H n \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n\right)\right]} \\
 &= \lim_{\rho \rightarrow +\infty} \frac{HH! \frac{S^S}{S!} \rho^H}{S\rho\mu\left[\sum_{n=0}^S (H-n) \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^{H-1} (H-n) \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n\right]} \\
 &= \lim_{\rho \rightarrow +\infty} \frac{HH! \frac{S^S}{S!} \rho^{H-1}}{S\mu\left[\frac{H!}{1!} \frac{S^S}{S!} \rho^{H-1}\right]} \\
 &= \frac{H}{S\mu}.
 \end{aligned}$$

De

$$\begin{aligned} W_q &= W_s - \frac{1}{\mu} \\ &= \frac{L_s}{S\rho\mu(H - L_s)} - \frac{1}{\mu}, \end{aligned}$$

supondo que $S = 10$, $H = 30$, obtemos os gráficos da Figura 47.

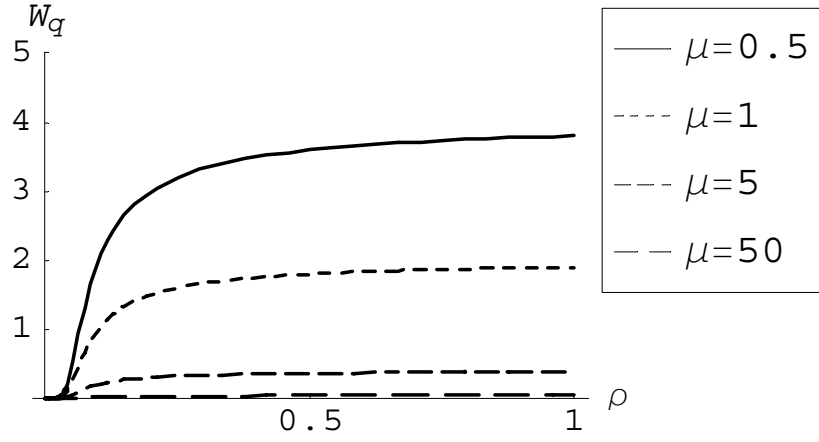


Figura 47: Tempo médio de espera na fila (W_q) em função da taxa de pressão (ρ), com $S = 10$ e $H = 30$

O gráfico da Figura 47 difere pouco da Figura 46, pois neste gráfico não estamos a contar com o tempo de espera que os clientes demoram a ser atendidos. Relativamente ao comportamento de W_q quando $\rho \rightarrow +\infty$, temos

$$\begin{aligned} \lim_{\rho \rightarrow +\infty} W_q &= \lim_{\rho \rightarrow +\infty} \left[W_s - \frac{1}{\mu} \right] = \frac{H}{S\mu} - \frac{1}{\mu} \\ &= \frac{H - S}{S\mu}. \end{aligned}$$

E, finalmente, o número médio de clientes na fila é dado por

$$L_q = L_s - \rho S (H - L_s),$$

e, quando ρ tende para $+\infty$, obtemos

$$\begin{aligned}
\lim_{\rho \rightarrow +\infty} L_q &= \lim_{\rho \rightarrow +\infty} [L_s - \rho S (H - L_s)] = H - S \lim_{\rho \rightarrow +\infty} \rho (H - L_s) \\
&= H - S \lim_{\rho \rightarrow +\infty} \rho \left(H - \sum_{n=0}^H n P_n \right) \\
&= H - S \lim_{\rho \rightarrow +\infty} \rho \left[H - P_0 \left(\sum_{n=0}^S n \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H n \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n \right) \right] \\
&= H - S \lim_{\rho \rightarrow +\infty} \rho \left[H - \frac{\sum_{n=0}^S n \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H n \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n}{\sum_{n=0}^S \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n} \right] \\
&= H - S \lim_{\rho \rightarrow +\infty} \rho \left[\frac{\sum_{n=0}^S (H-n) \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^{H-1} (H-n) \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n}{\frac{H!}{0!} \frac{S^S}{S!} \rho^H} \right] \\
&= H - S \lim_{\rho \rightarrow +\infty} \rho \left(\frac{\frac{H!}{1!} \frac{S^S}{S!} \rho^{H-1}}{H! \frac{S^S}{S!} \rho^H} \right) \\
&= H - S.
\end{aligned}$$

Pode representar graficamente na Figura 48, supondo que $S = 10$, $H = 30$, L_s e L_q em simultâneo.

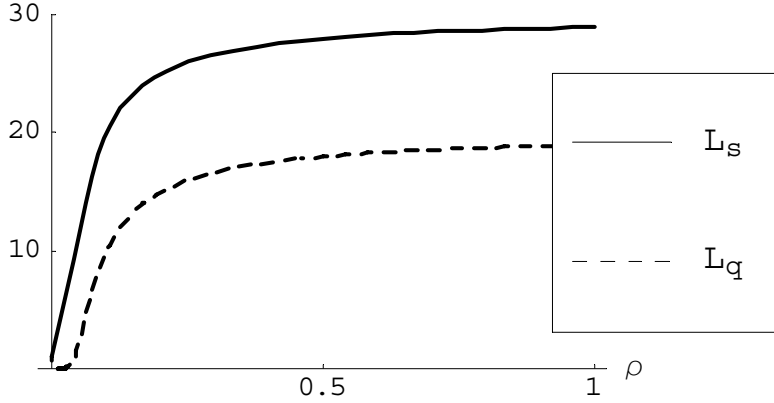


Figura 48: Número médio de clientes no sistema (L_s) e na fila (L_q) em função da taxa de pressão (ρ), com $S = 10$ e $H = 30$

Na Figura 48, observamos que o número de clientes aumenta consoante aumenta a taxa de pressão sem nunca atingir H . De facto, como já foi visto, L_s converge para H , e L_q converge para $H - S$.

Podemos resumir estas características no seguinte quadro,

Probabilidade de ocorrência do estado 0	$P_0 = \left[\sum_{n=0}^S \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} + \sum_{n=S+1}^H \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n \right]^{-1}$
Probabilidade de ocorrência do estado n	$P_n = \begin{cases} \frac{H!}{(H-n)!} \frac{(S\rho)^n}{n!} P_0, & n = 1, \dots, S \\ \frac{H!}{(H-n)!} \frac{S^S}{S!} \rho^n P_0, & n = S+1, \dots, H \\ 0, & n > H \end{cases}$
Número médio de clientes no sistema	$L_s = \sum_{n=0}^H n P_n$
Comprimento médio da fila	$L_q = L_s - \rho S (H - L_s)$
Tempo médio de espera no sistema	$W_s = \frac{L_s}{S\rho\mu(H-L_s)}$
Tempo médio de espera na fila	$W_q = \frac{L_q}{S\rho\mu(H-L_s)} - \frac{1}{\mu}$
Taxa de pressão	$\rho = \frac{\lambda}{S\mu}$

Quadro 8: Sumário das características do modelo $M/M/S/\infty/H$

5 Outros modelos

Na teoria das filas de espera existem variadíssimos modelos, muitos deles já estão estudados, mas ainda existem muitos a serem estudados, e há claro os modelos que ainda estão à espera de ser “inventados” e estudados.

Aqui já estudamos os modelos em que as chegadas são distribuídas segundo um processo de Poisson e o serviço de acordo com a distribuição exponencial.

Agora neste capítulo vamos tratar de alguns outros modelos que não seguem estas condições e que, por tal razão, por vezes não vão ser tratados com tanto pormenor, dada a sua complexidade.

5.1 Modelo $M/E_r/1$

O modelo que se segue contém um só servidor, as chegadas têm distribuição de Poisson com parâmetro λ e os tempos de serviço ocorrem com uma distribuição de Erlang- r com parâmetros r e μ . Por exemplo, este modelo aplica-se no caso em que temos um trabalho que tem de passar, etapa para etapa, através de uma série de r fases de produção independentes, e cada etapa tem um tempo com distribuição exponencial com um parâmetro comum μ . A análise do modelo $M/E_r/1$ não difere muito da do modelo $M/M/1$.

A taxa de ocupação é dada por $\rho = \frac{\lambda r}{\mu}$ e para que o sistema atinja o estado de equilíbrio é necessário e suficiente que $\rho < 1$ (cf. [4] e [3]).

Este modelo pode ser representado pela Figura 49.

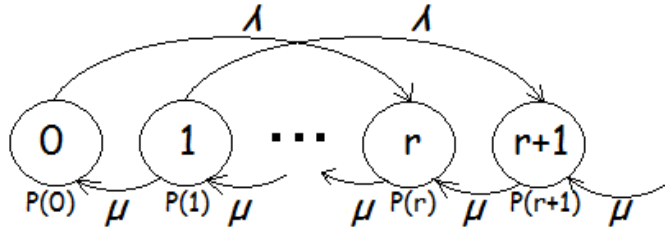


Figura 49: Modelo $M/E_r/1$

Vamos tentar determinar P_n , a probabilidade de existirem n clientes no sistema, equiparando a taxa de saída do estado n e a taxa de entrada no estado n :

$$\begin{cases} P_0 \lambda = P_1 \mu \\ P_n (\lambda + \mu) = P_{n+1} \mu, \quad n = 1, \dots, r-1 \\ P_n (\lambda + \mu) = P_{n-1} \lambda + P_{n+1} \mu, \quad n = r, \dots \end{cases} \quad (32)$$

Seja $P(z)$ a f.g.p. de P_n , ou seja

$$P(z) = \sum_{n=0}^{\infty} P_n z^n,$$

com $|z| \leq 1$.

Se multiplicarmos as expressões (32) por z^n , obtemos

$$\begin{cases} P_0 \lambda z^0 = P_1 \mu z^0 \\ P_n z^n (\lambda + \mu) = P_{n+1} z^n \mu, n = 1, \dots, r-1 \\ P_n z^n (\lambda + \mu) = P_{n-1} z^n \lambda + P_{n+1} z^n \mu, n = r, \dots \end{cases}$$

e, fazendo a adição de todos os estados n , obtemos

$$\begin{cases} P_0 \lambda z^0 = P_1 \mu z^0 \\ \sum_{n=1}^{r-1} P_n z^n (\lambda + \mu) = \sum_{n=1}^{r-1} P_{n+1} z^n \mu \\ \sum_{n=r}^{\infty} P_n z^n (\lambda + \mu) = \sum_{n=r}^{\infty} P_{n-1} z^n \lambda + \sum_{n=r}^{\infty} P_{n+1} z^n \mu \end{cases}.$$

Somando as expressões,

$$P_0 \lambda z^0 + \sum_{n=1}^{r-1} P_n z^n (\lambda + \mu) + \sum_{n=r}^{\infty} P_n z^n (\lambda + \mu) = P_1 \mu z^0 + \sum_{n=1}^{r-1} P_{n+1} z^n \mu + \sum_{n=r}^{\infty} P_{n-1} z^n \lambda + \sum_{n=r}^{\infty} P_{n+1} z^n \mu$$

$$\Leftrightarrow \sum_{n=0}^{\infty} P_n z^n (\lambda + \mu) - P_0 \mu = \frac{1}{z} \sum_{n=0}^{\infty} P_{n+1} z^{n+1} \mu + \sum_{n=r}^{\infty} P_{n-1} z^n \lambda$$

$$\Leftrightarrow P(z) (\lambda + \mu) - P_0 \mu = \frac{\mu}{z} \sum_{n=1}^{\infty} P_n z^n + \sum_{n=0}^{\infty} P_n z^{n+r} \lambda$$

$$\Leftrightarrow P(z) (\lambda + \mu) - P_0 \mu = \frac{\mu}{z} \left(\sum_{n=0}^{\infty} P_n z^n - P_0 z^0 \right) + \lambda z^r P(z)$$

$$\Leftrightarrow P(z) (\lambda + \mu) - P_0 \mu = \frac{\mu}{z} (P(z) - P_0) + \lambda z^r P(z).$$

Vamos agora escrever a expressão anterior em função de $P(z)$

$$P(z) (\lambda + \mu) - P_0 \mu = \frac{\mu}{z} (P(z) - P_0) + \lambda z^r P(z)$$

$$\Leftrightarrow P(z) \lambda z + P(z) \mu z - P_0 \mu z = P(z) \mu - P_0 \mu + \lambda z^{r+1} P(z)$$

$$\Leftrightarrow P(z) (\lambda z + \mu z - \mu - \lambda z^{r+1}) = P_0 \mu (z - 1)$$

$$\Leftrightarrow P(z) = \frac{P_0 \mu (z - 1)}{\lambda z + \mu (z - 1) - \lambda z^{r+1}} = \frac{P_0 \mu}{\frac{\lambda z(1 - z^r)}{z - 1} + \mu}$$

$$= \frac{P_0 \mu}{\mu - \lambda z \frac{1 - z^r}{1 - z}} = \frac{P_0 \mu}{\mu - \lambda z (1 + z + \dots + z^{r-1})}$$

$$= \frac{-P_0 \mu / \lambda}{z + \dots + z^r - \mu / \lambda}.$$

Sejam z_1, \dots, z_r os zeros de polinómio $z + \dots + z^r - \mu / \lambda$ e suponhamos que são todos distintos. Temos então

$$P(z) = \frac{-P_0 \mu / \lambda}{(z - z_1) \dots (z - z_r)} = \frac{-P_0 \mu / \lambda}{\prod_{i=1}^r (z - z_i)}.$$

Decompondo $P(z)$ na soma de fracção com denominador do tipo $z - z_k$, obtemos

$$P(z) = \sum_{k=1}^r \frac{A_k}{z - z_k},$$

onde $A_k = \frac{-P_0\mu/\lambda}{\prod_{i=1, i \neq k}^r (z_k - z_i)}$.

Logo

$$P(z) = \sum_{k=1}^r \frac{-P_0\mu/\lambda}{\prod_{i=1, i \neq k}^r (z_k - z_i)} \frac{1}{z - z_k}.$$

Derivando temos

$$\frac{d^n P(z)}{dz^n} = \sum_{k=1}^r \frac{-P_0\mu/\lambda}{\prod_{i=1, i \neq k}^r (z_k - z_i)} (-1)^n \frac{n!}{(z - z_k)^{n+1}},$$

e tomando $z = 0$ vem

$$\left. \frac{d^n P(z)}{dz^n} \right|_{z=0} = \sum_{k=1}^r \frac{P_0\mu/\lambda}{\prod_{i=1, i \neq k}^r (z_k - z_i)} \frac{n!}{z_k^{n+1}}.$$

Então temos que

$$\begin{aligned} P[X = n] &= \frac{1}{n!} \sum_{k=1}^r \frac{P_0\mu/\lambda}{\prod_{i=1, i \neq k}^r (z_k - z_i)} \frac{n!}{z_k^{n+1}} \\ &= \sum_{k=1}^r \frac{P_0\mu/\lambda}{z_k \prod_{i=1, i \neq k}^r (z_k - z_i)} \left(\frac{1}{z_k} \right)^n \\ &= \sum_{k=1}^r c_k \left(\frac{1}{z_k} \right)^n, \end{aligned}$$

onde $c_k = \frac{P_0\mu/\lambda}{z_k \prod_{i=1, i \neq k}^r (z_k - z_i)}$.

Vamos agora determinar P_0 . De

$$\begin{aligned} P_n &= \sum_{k=1}^r \frac{P_0\mu/\lambda}{z_k \prod_{i=1, i \neq k}^r [z_k (1 - z_i/z_k)]} \left(\frac{1}{z_k} \right)^n \\ &= \sum_{k=1}^r \frac{P_0\mu/\lambda}{z_k z_k^{r-1} \prod_{i=1, i \neq k}^r (1 - z_i/z_k)} \left(\frac{1}{z_k} \right)^n \\ &= P_0 \sum_{k=1}^r \frac{\mu/\lambda}{z_k^r \prod_{i=1, i \neq k}^r (1 - z_i/z_k)} \left(\frac{1}{z_k} \right)^n. \end{aligned}$$

e de $\sum_{n=0}^{\infty} P_n = 1$, temos que

$$\sum_{n=0}^{\infty} P_0 \sum_{k=1}^r \frac{\mu/\lambda}{z_k^r \prod_{i=1, i \neq k}^r (1 - z_i/z_k)} \left(\frac{1}{z_k}\right)^n = 1,$$

assim temos que P_0 é

$$P_0 = \left[\sum_{n=0}^{\infty} \sum_{k=1}^r \frac{\mu/\lambda}{z_k^r \prod_{i=1, i \neq k}^r (1 - z_i/z_k)} \left(\frac{1}{z_k}\right)^n \right]^{-1}.$$

Simplificando vem P_0

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{\infty} \sum_{k=1}^r \frac{\mu/\lambda}{z_k^r \prod_{i=1, i \neq k}^r (1 - z_i/z_k)} \left(\frac{1}{z_k}\right)^n \right]^{-1} \\ &= \left[\sum_{k=1}^r \frac{\mu/\lambda}{z_k^r \prod_{i=1, i \neq k}^r (1 - z_i/z_k)} \sum_{n=0}^{\infty} \left(\frac{1}{z_k}\right)^n \right]^{-1}. \end{aligned}$$

Se $\left|\frac{1}{z_k}\right| < 1$,

$$\begin{aligned} P_0 &= \left[\sum_{k=1}^r \frac{\mu/\lambda}{z_k^r \prod_{i=1, i \neq k}^r (1 - z_i/z_k)} \frac{1}{1 - 1/z_k} \right]^{-1} \\ &= \left[\sum_{k=1}^r \frac{\mu/\lambda}{z_k^{r-1} \prod_{i=1, i \neq k}^r (1 - z_i/z_k)} \frac{1}{z_k - 1} \right]^{-1}. \end{aligned}$$

Temos, segundo [2], que o tempo médio esperado pelo cliente na fila é

$$W_q = L_q \frac{r}{\mu} + \rho R \quad (33)$$

onde R corresponde ao valor esperado do tempo residual (T_R) do serviço do cliente que está a ser servido, que pode ser determinado da seguinte forma

$$\begin{aligned} E[T_R] &= \sum_{i=1}^r E[T_R | etapa i] P[etapa i] \\ &= \sum_{i=1}^r \frac{r - (i - 1)}{\mu} \frac{1}{r} = \frac{1}{r\mu} (r + r - 1 + \dots + 2 + 1) \\ &= \frac{r + 1}{2\mu}. \end{aligned}$$

Substituindo R na expressão (33) temos

$$W_q = L_q \frac{r}{\mu} + \rho \frac{r+1}{2\mu}.$$

Usando a formula de Little,

$$L_q = \lambda W_q,$$

resulta que o tempo médio que um cliente tem de esperar na fila é

$$\begin{aligned} W_q &= \lambda W_q \frac{r}{\mu} + \rho \frac{r+1}{2\mu} \\ &= \rho W_q + \rho \frac{r+1}{2\mu} \\ \Leftrightarrow W_q - \rho W_q &= \rho \frac{r+1}{2\mu} \\ \Leftrightarrow W_q (1 - \rho) &= \rho \frac{r+1}{2\mu} \\ \Leftrightarrow W_q &= \frac{\rho(r+1)}{2\mu(1-\rho)}. \end{aligned}$$

E assim também podemos determinar o número médio de clientes na fila

$$\begin{aligned} L_q &= \lambda W_q = \lambda \frac{\rho(r+1)}{2\mu(1-\rho)} \\ &= \frac{\rho^2(r+1)}{2r(1-\rho)}. \end{aligned}$$

E por último temos que o tempo médio que um cliente tem de esperar no sistema é a soma do tempo médio de um cliente na fila com o tempo médio que demora a ser atendido, ou seja,

$$\begin{aligned} W_s &= W_q + \frac{r}{\mu} = \frac{\rho(r+1)}{2\mu(1-\rho)} + \frac{r}{\mu} \\ &= \frac{\rho(r+1) + 2r(1-\rho)}{2\mu(1-\rho)} \\ &= \frac{\rho r + \rho + 2r - 2r\rho}{2\mu(1-\rho)} = \frac{\rho + 2r - \rho r}{2\mu(1-\rho)} \\ &= \frac{\rho + r(2-\rho)}{2\mu(1-\rho)}. \end{aligned}$$

Com o conhecimento do tempo médio de um cliente no sistema, é possível saber o número médio de clientes no sistema, usando as fórmulas de Little,

$$\begin{aligned} L_s &= \lambda W_s = \lambda \frac{\rho + r(2-\rho)}{2\mu(1-\rho)} \\ &= \rho \frac{\rho + r(2-\rho)}{2r(1-\rho)}. \end{aligned}$$

O seguinte quadro reúne as principais características do modelo $M/E_r/1$,

Probabilidade de ocorrência do estado 0	$P_0 = \left[\sum_{k=1}^r \frac{-\mu/\lambda}{z_k^r \prod_{i=1, i \neq k}^r (1-z_i/z_k)} \frac{z_k}{z_k-1} \right]^{-1}$
Probabilidade de ocorrência do estado n	$P_n = P_0 \sum_{k=1}^r \frac{-\mu/\lambda}{z_k^r \prod_{i=1, i \neq k}^r (1-z_i/z_k)} \left(\frac{1}{z_k} \right)^n$
Número médio de clientes no sistema	$L_s = \rho \frac{\rho+r(2-\rho)}{2r(1-\rho)}$
Comprimento médio da fila	$L_q = \frac{\rho^2(r+1)}{2r(1-\rho)}$
Tempo médio de espera no sistema	$W_s = \frac{\rho+r(2-\rho)}{2\mu(1-\rho)}$
Tempo médio de espera na fila	$W_q = \frac{\rho(r+1)}{2\mu(1-\rho)}$
Taxa de ocupação	$\rho = \frac{\lambda r}{\mu}$

Quadro 9: Sumário das características do modelo $M/E_r/1$

5.2 Modelo $M/G/1$

Consideremos um sistema em que as chegadas formam um processo de Poisson com parâmetro λ e o serviço tem uma distribuição generalizada. Suponhamos que os tempos de serviço são iid com valor médio $\frac{1}{\mu}$ e variância σ^2 . Denotemos por U o tempo de serviço para o n -ésimo cliente e por G a respectiva função de distribuição cumulativa, isto é,

$$G(t) = P[U \leq t].$$

Suponhamos que

$$E[U] = \frac{1}{\mu}$$

e que

$$\text{var}[U] = \sigma^2.$$

e considerando a transformada de Laplace de G :

$$\tilde{G}(a) = E[e^{-aU}] = \int_{t=0}^{\infty} e^{-at} dG(t).$$

Suponhamos ainda que existe apenas um servidor e os clientes são servidos de acordo com a disciplina FCFS.

Seja X_n o número de clientes no sistema imediatamente após a partida do n -ésimo cliente. Temos que $\{X_n, n \geq 0\}$ é uma cadeia de Markov a tempo discreto.

Temos de verificar quais são as condições para que exista um estado de equilíbrio para este modelo. Usando o facto de as probabilidades de transição serem dadas por:

$$P[X_{n+1} = k | X_n = i] = p_{jk} = \begin{cases} a_k & \text{se } j = 0, k \geq 0 \\ a_{k-j+1} & \text{se } k \geq j-1 \geq 0 \\ 0 & \text{se } j \geq 1, 0 \leq k < j-1 \end{cases} \quad (34)$$

onde a_k é a probabilidade de chegarem k clientes durante o tempo de serviço, ou seja,

$$a_k = P[\text{"}k\text{" clientes que surgem durante o tempo de serviço, } U].$$

Conclui-se assim que a matriz de transição de $\{X_n\}_{n \in \mathbb{N}}$ é dada por

$$P = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & . & . & . \\ a_0 & a_1 & a_2 & a_3 & . & . & . \\ 0 & a_0 & a_1 & a_2 & . & . & . \\ 0 & 0 & a_0 & a_1 & . & . & . \\ 0 & 0 & 0 & a_0 & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \end{bmatrix}$$

Devido ao facto das chegadas ocorrerem de acordo uma distribuição de Poisson com parâmetro λ , e usando o Teorema da Probabilidade Total, temos que

$$\begin{aligned} a_k &= \int_0^{+\infty} P[\text{"}k\text{" clientes que surgem durante } U | U = t] f_U(t) dt \\ &= \int_0^{+\infty} \frac{e^{-\lambda t}}{k!} (\lambda t)^k dG(t), \end{aligned} \quad (35)$$

provando assim que a cadeia é irredutível e aperiódica.

Consideremos

$$r_k = 1 - a_0 - \dots - a_k,$$

e

$$r = r_0 + r_1 + \dots = \sum_{k=0}^{\infty} r_k.$$

Prova-se que a cadeia de Markov é recorrente não-nula se e só se

$$r < 1.$$

De facto, a equação

$$\begin{aligned} P_n &= \sum_{j=0}^{\infty} P_j p_{jn} \\ &= P_0 a_n + \sum_{j=1}^{n+1} P_j a_{n-j+1}, \quad n = 0, 1, \dots \end{aligned} \quad (36)$$

tem solução se e só se $r < 1$.

Para cada n , adicionamos as equações para P_0, \dots, P_n , da equação (36), lado a lado. Resulta

$$\begin{cases} P_0 = P_0 a_0 + P_1 a_0 \\ P_0 + P_1 = P_0 a_0 + P_1 a_0 + P_0 a_1 + P_1 a_1 + P_2 a_0 \\ P_0 + P_1 + P_2 = P_0 a_0 + P_1 a_0 + P_0 a_1 + P_1 a_1 + P_2 a_0 + P_0 a_2 + P_1 a_2 + P_2 a_1 + P_3 a_0 \\ \dots \end{cases}$$

$$\begin{aligned}
&\Leftrightarrow \begin{cases} P_1 a_0 = P_0 (1 - a_0) \\ P_2 a_0 = P_0 (1 - a_0 - a_1) + P_1 (1 - a_0 - a_1) \\ P_3 a_0 = P_0 (1 - a_0 - a_1 - a_2) + P_1 (1 - a_0 - a_1 - a_2) + P_2 (1 - a_0 - a_1) \\ \dots \end{cases} \\
&\Leftrightarrow \begin{cases} P_1 a_0 = P_0 r_0 \\ P_2 a_0 = P_0 r_1 + P_1 r_1 \\ P_3 a_0 = P_0 r_2 + P_1 r_2 + P_2 r_1 \\ \dots \end{cases} .
\end{aligned} \tag{37}$$

É claro que para qualquer $P_0 \geq 0$ fixo, este sistema de equações tem uma única solução não-negativa. Assim, (36) tem uma solução não-negativa.

Somando todas as equações (37), notando que a $a_0 = 1 - r_0$ e que a $r = r_0 + r_1 + \dots$, temos

$$(1 - r_0) \sum_{n=1}^{\infty} P_n = P_0 r + (r - r_0) \sum_{n=1}^{\infty} P_n. \tag{38}$$

Se $r < 1$,

$$\begin{aligned}
&\Leftrightarrow (1 - r_0 - r + r_0) \sum_{n=1}^{\infty} P_n = P_0 r \\
&\Leftrightarrow \sum_{n=1}^{\infty} P_n = \frac{P_0 r}{1 - r} < \infty,
\end{aligned}$$

como sabemos que

$$\sum_{n=0}^{\infty} P_n = 1,$$

temos

$$\begin{aligned}
\sum_{n=1}^{\infty} P_n &= \sum_{n=0}^{\infty} P_n - P_0 \\
&\Leftrightarrow \frac{P_0 r}{1 - r} = 1 - P_0 \\
&\Leftrightarrow P_0 r = (1 - P_0) (1 - r) \\
&\Leftrightarrow P_0 r = 1 - r - P_0 + P_0 r \\
&\Leftrightarrow P_0 = 1 - r.
\end{aligned} \tag{39}$$

Portanto, se $r < 1$, todos os estados são recorrentes não-nulos aperiódicos.

Por outro lado, se $r \geq 1$, temos de (38), que $P_0 = 0$ (que implica $P_1 = P_2 = \dots = 0$) ou $P_0 > 0$ e $\sum_{n=1}^{\infty} P_n = \infty$. Consequentemente, se $r \geq 1$, os estados são não recorrentes não nulos.

Multiplicando a expressão (36) por z^n e fazendo o somatório para $n = 0, 1, \dots$, obtemos

$$\begin{aligned}
P(z) &= \sum_{n=0}^{\infty} P_n z^n = \sum_{n=0}^{\infty} z^n P_0 a_n + \sum_{n=0}^{\infty} z^n \sum_{j=1}^{n+1} P_j a_{n-j+1} \\
&= P_0 \sum_{n=0}^{\infty} a_n z^n + \sum_{j=1}^{\infty} P_j z^{j-1} \sum_{n=j-1}^{\infty} a_{n-j+1} z^{n-j+1}.
\end{aligned} \tag{40}$$

Usando (35) temos que

$$\begin{aligned}
 \sum_{n=0}^{\infty} a_n z^n &= \sum_{n=0}^{\infty} \left[\int_0^{+\infty} \frac{e^{-\lambda t}}{n!} (\lambda t)^n dG(t) \right] z^n \\
 &= \int_0^{+\infty} e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t z)^n}{n!} dG(t) \\
 &= \int_0^{+\infty} e^{-\lambda t} e^{\lambda t z} dG(t) = \int_0^{+\infty} e^{-\lambda t(1-z)} dG(t) \\
 &= \tilde{G}(\lambda(1-z)).
 \end{aligned}$$

Relativamente ao segundo somatório,

$$\begin{aligned}
 \sum_{j=1}^{\infty} P_j z^{j-1} &= \frac{1}{z} \left(\sum_{j=1}^{\infty} P_j z^j + P_0 z^0 - P_0 z^0 \right) \\
 &= \frac{1}{z} \left(\sum_{j=0}^{\infty} P_j z^j - P_0 \right) \\
 &= \frac{1}{z} (P(z) - P_0).
 \end{aligned}$$

Fazendo uma mudança de variável no terceiro somatório de (40) temos

$$\sum_{n=j-1}^{\infty} a_{n-j+1} z^{n-j+1} = \sum_{n=0}^{\infty} a_n z^n.$$

Assim já podemos simplificar a expressão de $P(z)$ dada por (40)

$$P(z) = P_0 \tilde{G}(\lambda(1-z)) + \frac{P(z) - P_0}{z} \tilde{G}(\lambda(1-z)).$$

Resolvendo em função de $P(z)$, temos

$$\begin{aligned}
 P(z) &= \frac{z P_0 \tilde{G}(\lambda - \lambda z) + P(z) \tilde{G}(\lambda - \lambda z) - P_0 \tilde{G}(\lambda - \lambda z)}{z} \\
 \Leftrightarrow z P(z) - P(z) \tilde{G}(\lambda - \lambda z) &= z P_0 \tilde{G}(\lambda - \lambda z) - P_0 \tilde{G}(\lambda - \lambda z) \\
 \Leftrightarrow P(z) \left(\tilde{G}(\lambda - \lambda z) - z \right) &= P_0 \tilde{G}(\lambda - \lambda z) (1 - z) \\
 \Leftrightarrow P(z) &= \frac{P_0 \tilde{G}(\lambda - \lambda z) (1 - z)}{\tilde{G}(\lambda - \lambda z) - z}.
 \end{aligned}$$

De (39) sabemos que

$$P_0 = 1 - r.$$

Logo

$$P_n = \frac{P^{(n)}(0)}{n!},$$

com

$$P(z) = \frac{(1-r)(1-z) \tilde{G}(\lambda - \lambda z)}{\tilde{G}(\lambda - \lambda z) - z}. \quad (41)$$

A expressão anterior é uma das formas da equação de transformação de Pollaczek-Khinchin (cf. [41]).

Se $|z| < 1$,

$$\left. \frac{d}{dz} P(z) \right|_{z=1} = \sum_{n=0}^{\infty} n P_n z^{n-1} \Big|_{z=1} = \sum_{n=0}^{\infty} n P_n = L_s$$

donde resulta que

$$\left. \frac{d}{dz} P(z) \right|_{z=1} = \left. \frac{d}{dz} \frac{(1-r) \tilde{G}(\lambda - \lambda z)(1-z)}{\tilde{G}(\lambda - \lambda z) - z} \right|_{z=1},$$

i.e.,

$$\frac{\left. \frac{d}{dz} P(z) \right|_{z=1}}{1-r} = \frac{[\tilde{G}(\lambda - \lambda z) - z] \frac{d}{dz} [\tilde{G}(\lambda - \lambda z)(1-z)] - [\tilde{G}(\lambda - \lambda z)(1-z)] \frac{d}{dz} [\tilde{G}(\lambda - \lambda z) - z]}{[\tilde{G}(\lambda - \lambda z) - z]^2} \Big|_{z=1}.$$

Seja $g(z) = \tilde{G}(\lambda - \lambda z)$. Temos

$$\begin{aligned} \frac{\left. \frac{d}{dz} P(z) \right|_{z=1}}{1-r} &= \lim_{z \rightarrow 1} \frac{[g(z) - z] \left[(1-z) \frac{d}{dz} g(z) - g(z) \right] - [g(z)(1-z)] \left[\frac{d}{dz} g(z) - 1 \right]}{[g(z) - z]^2} \\ &= \lim_{z \rightarrow 1} \frac{[g(z) - z] (1-z) \frac{d}{dz} g(z) - g^2(z) - [g(z)(1-z)] \frac{d}{dz} g(z) + g(z)}{[g(z) - z]^2} \\ &= \lim_{z \rightarrow 1} \frac{(-z)(1-z) \frac{d}{dz} g(z) - g^2(z) + g(z)}{[g(z) - z]^2}. \end{aligned}$$

Ora,

$$g(z)|_{z=1} = \tilde{G}(\lambda - \lambda z)|_{z=1} = \tilde{G}(0)|_{z=1} = 1$$

e, conseqüentemente, o cálculo acima de $\lim_{z \rightarrow 1} \frac{d}{dz} P(z)$ leva-nos a uma indeterminação do tipo $\frac{0}{0}$. Usando da regra de L'Hôpital temos

$$\begin{aligned} \lim_{z \rightarrow 1} \frac{d}{dz} P(z) &= (1-r) \lim_{z \rightarrow 1} \frac{(2z-1) \frac{d}{dz} g(z) + (z^2-z) \frac{d^2}{dz^2} g(z) - 2g(z) \frac{d}{dz} g(z) + \frac{d}{dz} g(z)}{2[g(z)-z] \left[\frac{d}{dz} g(z) - 1 \right]} \\ &= (1-r) \lim_{z \rightarrow 1} \frac{2z \frac{d}{dz} g(z) + (z^2-z) \frac{d^2}{dz^2} g(z) - 2g(z) \frac{d}{dz} g(z)}{2[g(z)-z] \left[\frac{d}{dz} g(z) - 1 \right]} \\ &= (1-r) \lim_{z \rightarrow 1} \frac{2[z-g(z)] \frac{d}{dz} g(z) + z(z-1) \frac{d^2}{dz^2} g(z)}{2[g(z)-z] \left[\frac{d}{dz} g(z) - 1 \right]} \\ &= (1-r) \lim_{z \rightarrow 1} \left\{ \frac{\frac{d}{dz} g(z)}{\left[1 - \frac{d}{dz} g(z) \right]} + \frac{(z^2-z) \frac{d^2}{dz^2} g(z)}{2[g(z)-z] \left[\frac{d}{dz} g(z) - 1 \right]} \right\}. \end{aligned}$$

De

$$\frac{d}{dz} \tilde{G}(z) = \frac{d}{dz} \left[\int_0^{\infty} e^{-zt} dG(t) \right] = \int_0^{\infty} (-t) e^{-zt} dG(t)$$

resulta

$$\frac{d}{dz} g(z) = \lambda \int_0^{\infty} t e^{-(\lambda - \lambda z)t} dG(t) \Rightarrow \left. \frac{d}{dz} g(z) \right|_{z=1} = \lambda \int_0^{\infty} t dG(t) = \lambda E[U] = \frac{\lambda}{\mu},$$

$$\lim_{z \rightarrow 1} \frac{\frac{d}{dz}g(z)}{\left[1 - \frac{d}{dz}g(z)\right]} = \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}} = \frac{\lambda}{\mu - \lambda}$$

e

$$\begin{aligned} \frac{d^2}{dz^2}g(z) &= \frac{d}{dz} \left[\frac{d}{dz} \tilde{G}(\lambda - \lambda z) \right] \\ &= \frac{d}{dz} \lambda \int_0^\infty t e^{-(\lambda - \lambda z)t} dG(t) = \lambda^2 \int_0^\infty t^2 e^{-(\lambda - \lambda z)t} dG(t) \\ &\Rightarrow \left. \frac{d^2}{dz^2} \tilde{G}(\lambda - \lambda z) \right|_{z=1} = \lambda^2 \int_0^\infty t^2 dG(t) = \lambda^2 E[U^2]. \end{aligned}$$

Ora,

$$\begin{aligned} \lim_{z \rightarrow 1} \frac{(z^2 - z) \frac{d^2}{dz^2}g(z)}{2[g(z) - z] \left[\frac{d}{dz}g(z) - 1 \right]} &= \lim_{z \rightarrow 1} \frac{(2z - 1) \frac{d^2}{dz^2}g(z) + (z^2 - z) \frac{d^3}{dz^3}g(z)}{2 \left[\frac{d}{dz}g(z) - 1 \right] \left[\frac{d}{dz}g(z) - 1 \right] + 2[g(z) - z] \left[\frac{d^2}{dz^2}g(z) - 1 \right]} \\ &= \frac{\lambda^2 E[U^2]}{2 \left(\frac{\lambda}{\mu} - 1 \right)^2}. \end{aligned}$$

Substituindo em $\lim_{z \rightarrow 1} \frac{d}{dz}P(z)$ obtemos

$$\begin{aligned} L_s &= \lim_{z \rightarrow 1} \frac{d}{dz}P(z) = (1 - r) \lim_{z \rightarrow 1} \left\{ \frac{\frac{d}{dz}g(z)}{\left[1 - \frac{d}{dz}g(z)\right]} + \frac{(z^2 - z) \frac{d^2}{dz^2}g(z)}{2[g(z) - z] \left[\frac{d}{dz}g(z) - 1 \right]} \right\} \\ &= (1 - r) \left[\frac{\lambda}{\mu - \lambda} + \frac{\lambda^2 E[U^2]}{2 \left(\frac{\lambda}{\mu} - 1 \right)^2} \right] = (1 - r) \left[\rho \frac{1}{1 - \rho} + \frac{\lambda^2 E[U^2]}{2(1 - \rho)^2} \right] \\ &= \frac{1 - r}{1 - \rho} \left[\rho + \frac{\lambda^2 E[U^2]}{2(1 - \rho)} \right]. \end{aligned}$$

Considerando $E[U^2] = \sigma^2 + (E[U])^2 = \sigma^2 + \frac{1}{\mu^2}$ temos

$$\begin{aligned} L_s &= \frac{1 - r}{1 - \rho} \left[\rho + \frac{\lambda^2 \left(\sigma^2 + \frac{1}{\mu^2} \right)}{2(1 - \rho)} \right] = \frac{1 - r}{1 - \rho} \left[\rho + \frac{\lambda^2 \sigma^2 + \frac{\lambda^2}{\mu^2}}{2(1 - \rho)} \right] \\ &= \frac{1 - r}{1 - \rho} \left[\rho + \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} \right] = \frac{1 - r}{1 - \rho} \left[\frac{2\rho(1 - \rho) + \lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} \right] \\ &= \frac{1 - r}{1 - \rho} \left[\frac{2\rho - 2\rho^2 + \lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} \right] \\ &= \frac{1 - r}{1 - \rho} \left[\frac{2\rho - \rho^2 + \lambda^2 \sigma^2}{2(1 - \rho)} \right]. \end{aligned} \tag{42}$$

Usando a fórmula de Little obtemos

$$W_s = \frac{L_s}{\lambda} = \frac{1 - r}{1 - \rho} \left[\frac{2\rho - \rho^2 + \lambda^2 \sigma^2}{2\lambda(1 - \rho)} \right], \tag{43}$$

de onde resulta que

$$\begin{aligned}
 W_q &= W_s - \frac{1}{\mu} \\
 &= \frac{1-r}{1-\rho} \left[\frac{2\rho - \rho^2 + \lambda^2 \sigma^2}{2\lambda(1-\rho)} \right] - \frac{\rho}{\lambda} \\
 &= \frac{(1-r)(2\rho - \rho^2 + \lambda^2 \sigma^2) - 2\rho(1-\rho)^2}{2\lambda(1-\rho)^2} \\
 &= \frac{2\rho - 2\rho r - \rho^2 + \rho^2 r + \lambda^2 \sigma^2 - \lambda^2 \sigma^2 r - 2\rho(1-\rho)^2}{2\lambda(1-\rho)^2} \\
 &= \frac{2\rho(1-r) - \rho^2(1-r) + \lambda^2 \sigma^2(1-r) - 2\rho(1-\rho)^2}{2\lambda(1-\rho)^2} \\
 &= \frac{(2\rho - \rho^2 + \lambda^2 \sigma^2)(1-r) - 2\rho(1-\rho)^2}{2\lambda(1-\rho)^2}
 \end{aligned} \tag{44}$$

e que

$$\begin{aligned}
 L_q &= \lambda W_q = \lambda \frac{(2\rho - \rho^2 + \lambda^2 \sigma^2)(1-r) - 2\rho(1-\rho)^2}{2\lambda(1-\rho)^2} \\
 &= \frac{(2\rho - \rho^2 + \lambda^2 \sigma^2)(1-r) - 2\rho(1-\rho)^2}{2(1-\rho)^2}.
 \end{aligned} \tag{45}$$

Podemos verificar que substituindo a variância σ^2 por $\frac{1}{\mu^2}$, que é o que acontece quando o tempo de serviço é Exponencial(μ), o número médio de clientes no sistema ou na fila, e o tempo médio de espera de um cliente no sistema ou na fila correspondem aos do modelo $M/M/1$.

De facto, de (35) temos que

$$a_k = \int_0^{+\infty} \frac{e^{-\lambda t}}{k!} (\lambda t)^k dG(t) \tag{46}$$

e, quando o serviço é exponencial

$$G(t) = (1 - e^{-\mu t}) I_{(0,+\infty)}(t).$$

Substituindo $G(t)$ em (46) temos

$$\begin{aligned}
 a_k &= \int_0^{+\infty} \frac{e^{-\lambda t}}{k!} (\lambda t)^k \mu e^{-\mu t} dt \\
 &= \frac{\mu}{k!} \int_0^{+\infty} e^{-(\lambda+\mu)t} (\lambda t)^k dt.
 \end{aligned}$$

Fazendo a mudança de variável

$$v = (\lambda + \mu)t \Rightarrow t = \frac{v}{\lambda + \mu} \Rightarrow dt = \frac{1}{\lambda + \mu} dv,$$

obtemos

$$\begin{aligned}
 a_k &= \frac{\mu}{k!} \int_0^{+\infty} e^{-v} \lambda^k \left(\frac{v}{\lambda + \mu} \right)^k \frac{1}{\lambda + \mu} dv \\
 &= \frac{\mu}{k!} \left(\frac{1}{\lambda + \mu} \right)^{k+1} \lambda^k \int_0^{+\infty} e^{-v} v^k dv \\
 &= \frac{\mu}{k!} \left(\frac{1}{\lambda + \mu} \right)^{k+1} \lambda^k \Gamma(k+1) \\
 &= \frac{\mu}{k!} \frac{1}{\lambda + \mu} \left(\frac{\lambda}{\lambda + \mu} \right)^k k! \\
 &= \frac{\mu}{\lambda + \mu} \left(\frac{\lambda}{\lambda + \mu} \right)^k.
 \end{aligned}$$

Temos então que

$$\begin{aligned}
 r &= \sum_{k=0}^{\infty} r_k = \sum_{k=0}^{\infty} \left(1 - \sum_{i=0}^{\infty} a_i \right) = \sum_{k=0}^{\infty} \sum_{i=k+1}^{\infty} a_i \\
 &= \sum_{k=0}^{\infty} \sum_{i=k+1}^{\infty} \frac{\mu}{\lambda + \mu} \left(\frac{\lambda}{\lambda + \mu} \right)^i = \frac{\mu}{\lambda + \mu} \sum_{k=0}^{\infty} \left(\frac{\lambda}{\lambda + \mu} \right)^{k+1} \frac{1}{1 - \frac{\lambda}{\lambda + \mu}} \\
 &= \frac{\mu}{\lambda + \mu} \frac{\lambda + \mu}{\lambda + \mu - \lambda} \sum_{k=0}^{\infty} \left(\frac{\lambda}{\lambda + \mu} \right)^{k+1} = \frac{\lambda}{\lambda + \mu} \frac{1}{1 - \frac{\lambda}{\lambda + \mu}} \\
 &= \frac{\lambda}{\lambda + \mu} \frac{\lambda + \mu}{\mu} = \frac{\lambda}{\mu} = \rho.
 \end{aligned}$$

Substituindo em (44), (43), (42) e em (45) σ^2 por $\frac{1}{\mu^2}$ e r por ρ , obtemos o número médio de clientes no sistema ou na fila, e o tempo médio de espera de um cliente no sistema ou na fila do modelo $M/M/1$:

$$\begin{aligned}
 W_q &= \frac{\left(2\rho - \rho^2 + \lambda^2 \frac{1}{\mu^2} \right) (1 - \rho) - 2\rho (1 - \rho)^2}{2\lambda (1 - \rho)^2} \\
 &= \frac{2\rho - \rho^2 + \rho^2 - 2\rho (1 - \rho)}{2\lambda (1 - \rho)} = \frac{2\rho - 2\rho + 2\rho^2}{2\lambda (1 - \rho)} \\
 &= \frac{\rho^2}{\lambda (1 - \rho)} = \frac{\rho^2}{\rho\mu (1 - \rho)} \\
 &= \frac{\rho}{\mu (1 - \rho)},
 \end{aligned}$$

$$\begin{aligned}
 W_s &= \frac{2\rho - \rho^2 + \lambda^2 \frac{1}{\mu^2}}{2\lambda (1 - \rho)} = \frac{\rho}{\lambda (1 - \rho)} \\
 &= \frac{1}{\mu (1 - \rho)},
 \end{aligned}$$

$$\begin{aligned} L_s &= \frac{2\rho - \rho^2 + \lambda^2 \frac{1}{\mu^2}}{2(1-\rho)} \\ &= \frac{\rho}{1-\rho}, \end{aligned}$$

e, finalmente,

$$\begin{aligned} L_q &= \frac{\left(2\rho - \rho^2 + \lambda^2 \frac{1}{\mu^2}\right)(1-\rho) - 2\rho(1-\rho)^2}{2(1-\rho)^2} \\ &= \frac{2\rho - 2\rho + 2\rho^2}{2(1-\rho)} \\ &= \frac{\rho^2}{1-\rho}. \end{aligned}$$

O seguinte quadro reúne as principais características do modelo $M/G/1$:

Probabilidade de ocorrência do estado 0	$P_0 = 1 - r$
Probabilidade de ocorrência do estado n	$P_n = \frac{P^{(n)}(0)}{n!}$, P dado por (41).
Número médio de clientes no sistema	$L_s = \frac{1-r}{1-\rho} \left[\frac{2\rho - \rho^2 + \lambda^2 \sigma^2}{2(1-\rho)} \right]$
Comprimento médio da fila	$L_q = \frac{(2\rho - \rho^2 + \lambda^2 \sigma^2)(1-r) - 2\rho(1-\rho)^2}{2(1-\rho)^2}$
Tempo médio de espera no sistema	$W_s = \frac{1-r}{1-\rho} \left[\frac{2\rho - \rho^2 + \lambda^2 \sigma^2}{2\lambda(1-\rho)} \right]$
Tempo médio de espera na fila	$W_q = \frac{(2\rho - \rho^2 + \lambda^2 \sigma^2)(1-r) - 2\rho(1-\rho)^2}{2\lambda(1-\rho)^2}$
Taxa de ocupação	$\rho = \frac{\lambda}{\mu}$

Quadro 10: Sumário das características do modelo $M/G/1$

5.3 Modelo $M/G/\infty$

Embora este modelo seja uma extensão do modelo $M/M/\infty$, os resultados são idênticos, pois os resultados do modelo $M/M/\infty$ são independentes da distribuição do tempo de serviço. A probabilidade de existirem n clientes no sistema ao longo do tempo é, tendo em conta que $\rho = \frac{\lambda}{\mu}$,

$$P_n = \frac{\rho^n}{n!} e^{-\rho}, \quad n \geq 0.$$

O comprimento médio da fila, L_q , e do sistema, L_s , o tempo de espera médio da fila, W_q , e do sistema, W_s , correspondem assim ao do modelo $M/M/\infty$, ou seja,

$$\begin{aligned} L_s &= \rho, \\ W_s &= \frac{1}{\mu}, \end{aligned}$$

$$L_q = W_q = 0.$$

O seguinte quadro contém as principais características do modelo $M/G/\infty$,

Probabilidade de ocorrência do estado 0	$P_0 = e^{-\rho}$
Probabilidade de ocorrência do estado n	$P_n = \frac{\rho^n}{n!} e^{-\rho}$
Número médio de clientes no sistema	$L_s = \rho$
Comprimento médio da fila	$L_q = 0$
Tempo médio de espera no sistema	$W_s = \frac{1}{\mu}$
Tempo médio de espera na fila	$W_q = 0$
Taxa de ocupação	$\rho = \frac{\lambda}{\mu}$

Quadro 11: Sumário das características do modelo $M/G/\infty$

5.4 Modelo $M/G/S/S$

Este modelo corresponde a um sistema onde as chegadas dos clientes ocorrem de acordo com um processo de Poisson com média $\frac{1}{\lambda}$ e o serviço de acordo com uma distribuição generalizada com média $\frac{1}{\mu}$.

Supomos que existem S servidores e que não existe espaço para espera. Os resultados são idênticos aos resultados do modelo $M/M/S/S$. De facto, a fórmula de perda de Erlang foi desenvolvida inicialmente para este modelo, $M/G/S/S$ (cf. [35]).

Para $0 \leq n \leq S$ a probabilidade de existirem n clientes no sistema quando este está em equilíbrio é

$$P_n = \frac{(S\rho)^n / n!}{\sum_{k=0}^S (S\rho)^k / k!}.$$

A partida do processo é a fila com o processo de Poisson com parâmetro $(1 - P_S)\lambda$, ou seja, $\lambda' = (1 - P_S)\lambda$, o número médio de clientes no sistema é

$$L_s = S\rho(1 - P_S)$$

e o tempo de espera médio de um cliente no sistema é

$$\begin{aligned} W_s &= \frac{L_s}{\lambda'} = \frac{S\rho(1 - P_S)}{(1 - P_S)\lambda} \\ &= \frac{1}{\mu}. \end{aligned}$$

E, obviamente, o número médio de clientes na fila e o tempo médio de cliente à espera na fila é zero, ou seja,

$$L_q = W_q = 0.$$

O seguinte quadro reúne as principais características do modelo $M/G/S/S$,

Probabilidade de ocorrência do estado 0	$P_0 = \left[\sum_{k=0}^S (S\rho)^k / k! \right]^{-1}$
Probabilidade de ocorrência do estado n	$P_n = \frac{(S\rho)^n / n!}{\sum_{k=0}^S (S\rho)^k / k!}$
Número médio de clientes no sistema	$L_s = S\rho(1 - P_S)$
Comprimento médio da fila	$L_q = 0$
Tempo médio de espera no sistema	$W_s = \frac{1}{\mu}$
Tempo médio de espera na fila	$W_q = 0$
Taxa de ocupação	$\rho = \frac{\lambda}{S\mu}$

Quadro 12: Sumário das características do modelo $M/G/S/S$

5.5 Modelo $G/M/1$

A fila $G/M/1$, é digamos, o par da fila $M/G/1$ onde as chegadas ocorrem de acordo com um processo geral, mas os tempos de serviço têm uma distribuição exponencial. Consideremos o sistema onde o tempo entre as chegadas tem uma distribuição genérica e o tempo de serviço tem uma distribuição exponencial com valor médio $\frac{1}{\mu}$. Supomos ainda que o tempo entre as chegadas são iid com uma função de distribuição cumulativa G e valor médio $\frac{1}{\lambda}$ e que a sequência entre chegadas e os tempos de serviço são independentes. Temos que A_n é o instante da chegada do n -ésimo cliente ($n \geq 1$).

$$G(t) = \begin{cases} P[A_{n+1} - A_n \leq t], & \text{se } t > 0 \\ 0, & \text{se } t = 0 \end{cases}$$

e

$$E[A_{n+1} - A_n] = \int_0^{\infty} t dG(t) = \frac{1}{\lambda}.$$

Este modelo tem somente 1 servidor e os clientes seguem uma disciplina FCFS. Novamente, denotemos a transformada de Laplace de G por

$$\tilde{G}(a) = E \left[e^{-a(A_{n+1} - A_n)} \right] = \int_0^{\infty} e^{-at} dG(t).$$

Seja X_n o número de clientes no sistema após o n -ésimo cliente chegar ao sistema, $\{X_n, n \geq 0\}$. Se D_{k+1} for o número de clientes que são servidos entre a chegada do k -ésimo cliente e do $k+1$ -ésimo cliente, então

$$X_{n+1} = X_n + 1 - D_{n+1}.$$

Começemos por verificar quais são as condições para que exista um estado de equilíbrio neste modelo. Usando o facto de as probabilidades de transição serem dadas por

$$\begin{aligned}
p_{jk} &= \begin{cases} P[X_{n+1} = k | X_n = j], & \text{se } k \leq j+1 \\ 0, & \text{se } k > j+1 \end{cases} \\
&= \begin{cases} P[X_n + 1 - D_{n+1} = k | X_n = j], & \text{se } k \leq j+1 \\ 0, & \text{se } k > j+1 \end{cases} \\
&= \begin{cases} P[j+1 - D_{n+1} = k | X_n = j], & \text{se } k \leq j+1 \\ 0, & \text{se } k > j+1 \end{cases} \\
&= \begin{cases} P[D_{n+1} = j+1 - k | X_n = j], & \text{se } k \leq j+1 \\ 0, & \text{se } k > j+1 \end{cases}.
\end{aligned}$$

e as igualdades

$$P_k = \sum_{j=0}^{\infty} P_j p_{jk}, \quad k = 0, 1, \dots \quad (47)$$

e

$$\sum_{k=0}^{\infty} P_k = 1,$$

é possível mostrar que a condição para que o sistema esteja estabilizado é que $\rho = \frac{\lambda}{\mu} < 1$ (a prova é análoga à da fila $M/G/1$).

Seja

$$\begin{aligned}
b_n &= P \left[\begin{array}{l} \text{ocorrer } n \text{ partidas num intervalo de tempo entre chegadas} \\ \text{servidor está ocupado durante o tempo entre as chegadas} \end{array} \right] \\
&= P[D_{k+1} = n | \text{o servidor estar ocupado entre as chegadas}].
\end{aligned}$$

A matriz de transição de $\{X_n\}_{n \in \mathbb{N}}$ é então dada por

$$P = \begin{bmatrix} p_{00} & b_0 & 0 & 0 & . & . & . \\ p_{10} & b_1 & b_0 & 0 & . & . & . \\ p_{20} & b_2 & b_1 & b_0 & . & . & . \\ p_{30} & b_3 & b_2 & b_1 & . & . & . \\ p_{40} & b_4 & b_3 & b_2 & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \end{bmatrix}.$$

Temos

$$b_n = \int_0^{\infty} \frac{(\mu x)^n}{n!} e^{-\mu x} dG(x), \quad n = 0, 1, \dots$$

e

$$\sum_{n=0}^{\infty} b_n z^n = \int_0^{\infty} e^{-\mu x(1-z)} dG(x) = \tilde{G}(\mu(1-z)).$$

De (47) temos que

$$\begin{aligned}
 P_n &= \sum_{k=0}^{\infty} P_k p_{kn} = \sum_{k=n-1}^{\infty} P_k p_{kn} \\
 &= \sum_{k=0}^{\infty} P_{k+n-1} p_{k+n-1,n} \\
 &= \sum_{k=0}^{\infty} b_k P_{k+n-1}, n = 1, 2, \dots
 \end{aligned}$$

Quando $n = 0$,

$$\begin{aligned}
 P_0 &= \sum_{j=0}^{\infty} P_j p_{j0} = \sum_{j=1}^{\infty} P_j \left(1 - \sum_{k=0}^{\infty} b_k \right) \\
 &= \sum_{j=1}^{\infty} P_j \sum_{k=j+1}^{\infty} b_k.
 \end{aligned}$$

Se procurarmos solução da forma

$$P_n = \sigma^n, n = 0, 1, 2, \dots$$

resulta

$$\begin{aligned}
 \sigma^n &= \sum_{k=0}^{\infty} b_k \sigma^{k+n-1} \Rightarrow \sigma = \sum_{k=0}^{\infty} b_k \sigma^k \Rightarrow \\
 &\Rightarrow \sigma = \tilde{G}(\mu(1-\sigma)).
 \end{aligned}$$

Supondo $\rho < 1$, existe sempre uma única solução real para $\sigma = \tilde{G}(\mu(1-\sigma))$ no intervalo de $0 < \sigma < 1$, pois a função $f(\sigma) = \tilde{G}(\mu(1-\sigma))$ tem derivada crescente em $(0, 1)$, $f(0) > 0$ e $f'(1) > 1$, desde que $\rho < 1$.

A solução será então dada por

$$P_n = c\sigma^n$$

onde σ é a única raiz de $\sigma = \tilde{G}(\mu(1-\sigma))$ e c é uma constante de normalização cujo valor é facilmente retirado da igualdade $\sum_{n=0}^{\infty} P_n = 1$:

$$\begin{aligned}
 \sum_{n=0}^{\infty} P_n &= 1 \Rightarrow \sum_{n=0}^{\infty} c\sigma^n = 1 \\
 &\Rightarrow c = \frac{1}{\sum_{n=0}^{\infty} \sigma^n} = 1 - \sigma.
 \end{aligned}$$

Conclui-se assim que, quando o sistema está em situação de equilíbrio, o número de clientes no sistema tem distribuição Geométrica(σ):

$$P_n = (1 - \sigma) \sigma^n, n = 0, 1, 2, \dots$$

O tempo médio que um cliente espera no sistema é

$$W_s = \frac{1}{\mu} L_a + \frac{1}{\mu}$$

onde L_a é uma v.a. que indica o número médio de clientes no sistema encontrados no acto da chegada.

De

$$\begin{aligned} L_a &= \sum_{n=0}^{\infty} n P_n = \sum_{n=0}^{\infty} n (1 - \sigma) \sigma^n \\ &= \frac{\sigma}{1 - \sigma} \end{aligned}$$

resulta

$$\begin{aligned} W_s &= \frac{1}{\mu} \frac{\sigma}{1 - \sigma} + \frac{1}{\mu} = \frac{\sigma + 1 - \sigma}{\mu (1 - \sigma)} \\ &= \frac{1}{\mu (1 - \sigma)}. \end{aligned} \quad (48)$$

Usando as fórmulas de Little, obtemos o número médio de clientes no sistema:

$$\begin{aligned} L_s &= \lambda W_s = \lambda \frac{1}{\mu (1 - \sigma)} \\ &= \frac{\rho}{1 - \sigma}, \end{aligned} \quad (49)$$

o número médio de clientes na fila,

$$\begin{aligned} L_q &= L_s - \rho = \frac{\rho}{1 - \sigma} - \rho \\ &= \frac{\rho \sigma}{1 - \sigma} \end{aligned} \quad (50)$$

e o tempo médio que um cliente espera na fila

$$\begin{aligned} W_q &= W_s - \frac{1}{\mu} = \frac{1}{\mu (1 - \sigma)} - \frac{1}{\mu} \\ &= \frac{\sigma}{\mu (1 - \sigma)}. \end{aligned} \quad (51)$$

Quando as chegadas ocorrem através de um processo de Poisson obtemos o modelo $M/M/1$. Neste caso a solução da equação $\sigma = \tilde{G}(\mu(1 - \sigma))$ é dada por

$$\begin{aligned} \sigma &= \tilde{G}(\mu(1 - \sigma)) = \int_0^{\infty} e^{-\mu(1 - \sigma)t} dG(t) \\ &= \int_0^{\infty} e^{-\mu(1 - \sigma)t} \lambda e^{-\lambda t} dt = \lambda \int_0^{\infty} e^{-[\mu(1 - \sigma) + \lambda]t} dt \\ &= \lambda \left. \frac{e^{-[\mu(1 - \sigma) + \lambda]t}}{-[\mu(1 - \sigma) + \lambda]} \right|_0^{\infty} = \lambda \left\{ 0 - \frac{1}{-[\mu(1 - \sigma) + \lambda]} \right\} \\ &= \frac{\lambda}{\mu - \sigma\mu + \lambda}. \end{aligned}$$

Ora

$$\begin{aligned}
\sigma &= \frac{\lambda}{\mu - \sigma\mu + \lambda} \Leftrightarrow \sigma(\mu - \sigma\mu + \lambda) = \lambda \\
&\Leftrightarrow \sigma\mu - \sigma^2\mu + \sigma\lambda - \lambda = 0 \\
&\Leftrightarrow \sigma\mu(1 - \sigma) + \lambda(\sigma - 1) = 0 \\
&\Leftrightarrow (1 - \sigma)(\sigma\mu - \lambda) = 0 \\
&\Leftrightarrow 1 - \sigma = 0 \vee \sigma\mu - \lambda = 0 \\
&\Leftrightarrow \sigma = 1 \vee \sigma = \frac{\lambda}{\mu} = \rho.
\end{aligned}$$

Visto que $0 < \sigma < 1$, σ não pode ser igual a 1, por isso resta-nos que $\sigma = \rho$. Substituindo σ por ρ no número médio de clientes no sistema, (49), e na fila, (50), e no tempo médio de um cliente no sistema, (48), e na fila, (51), do modelo $M/M/1$ é

$$\begin{aligned}
L_s &= \frac{\rho}{1 - \rho}, \\
L_q &= \frac{\rho^2}{1 - \rho}, \\
W_s &= \frac{1}{\mu(1 - \rho)}, \\
W_q &= \frac{\rho}{\mu(1 - \rho)}.
\end{aligned}$$

O seguinte quadro reúne as principais características do modelo $G/M/1$,

Probabilidade de ocorrência do estado 0	$P_0 = 1 - \sigma$
Probabilidade de ocorrência do estado n	$P_n = (1 - \sigma) \sigma^n$
Número médio de clientes no sistema	$L_s = \frac{\rho}{1 - \sigma}$
Comprimento médio da fila	$L_q = \frac{\rho\sigma}{1 - \sigma}$
Tempo médio de espera no sistema	$W_s = \frac{1}{\mu(1 - \sigma)}$
Tempo médio de espera na fila	$W_q = \frac{\sigma}{\mu(1 - \sigma)}$
Taxa de ocupação	$\rho = \frac{\lambda}{\mu}$

Quadro 13: Sumário das características do modelo $G/M/1$

5.6 Modelo $G/G/1$

O sistema do modelo $G/G/1$ é composto de um único servidor com distribuição entre chegadas e tempos de serviço genéricos, e independentes.

Sejam O a variável que representa o tempo entre as chegadas, com média $\frac{1}{\lambda}$ e com variância σ_O^2 e U a variável que representa o tempo de serviço, com média $\frac{1}{\mu}$ e variância σ_U^2 .

Para este caso não existem resultados exactos, como nos casos anteriores, dada a complexidade do modelo, mas é possível encontrar valores aproximados para o número médio de clientes no sistema, L_s , na fila, L_q , e para o tempo médio de espera de um cliente no sistema, W_s e na fila, W_q .

Segundo Marchal [29] e Marshall [30], quando o sistema se encontra em equilíbrio, o tempo médio de espera de um cliente na fila é limitado da seguinte forma

$$\frac{\rho^2 (1 + C_U^2) - 2\rho}{2\lambda(1 - \rho)} \leq W_q \leq \frac{\lambda(\sigma_O^2 + \sigma_U^2)}{2(1 - \rho)},$$

isto é,

$$\frac{\rho(1 + C_U^2) - 2}{2\mu(1 - \rho)} \leq W_q \leq \frac{\rho\mu(\sigma_O^2 + \sigma_U^2)}{2(1 - \rho)}, \quad (52)$$

onde $C_U = \sigma_U/\mu$ e $\rho = \frac{\lambda}{\mu}$. A condição para que exista a estabilidade é $\rho < 1$.

O limite inferior de (52) não é muito bom, devido ao facto de tomar valores negativos na grande maioria dos casos, a menos que C_U seja maior que 1, isto é, que o serviço seja “mais aleatório” do que a exponencial que tem $C_U = 1$. De facto, para que o limite inferior seja maior do que zero é necessário que

$$\frac{\rho^2 (1 + C_U^2) - 2\rho}{2\lambda(1 - \rho)} > 0 \Rightarrow \rho^2 (1 + C_U^2) - 2\rho > 0$$

$$\Rightarrow \rho(1 + C_U^2) > 2 \Rightarrow 1 + C_U^2 > \frac{2}{\rho}$$

$$\Rightarrow C_U^2 > \frac{2}{\rho} - 1.$$

Como

$$0 < \rho < 1 \Rightarrow \frac{1}{\rho} > 1 \Rightarrow \frac{2}{\rho} > 2$$

vem

$$\begin{aligned} C_U^2 &> 2 - 1 \\ C_U^2 &> 1 \Rightarrow C_U > 1. \end{aligned}$$

Usando as fórmulas de Little é possível encontrar valores que limitam L_q , W_s e L_s . Relativamente ao número médio de clientes na fila sabemos que

$$L_q = \lambda W_q$$

logo

$$\frac{\rho^2 (1 + C_U^2) - 2\rho}{2(1 - \rho)} \leq L_q \leq \frac{(\rho\mu)^2 (\sigma_O^2 + \sigma_U^2)}{2(1 - \rho)}.$$

Quanto ao tempo médio de espera de um cliente no sistema, de

$$W_s = W_q + \frac{1}{\mu}$$

resulta que

$$\frac{\rho(1 + C_U^2) - 2}{2\mu(1 - \rho)} + \frac{1}{\mu} \leq W_s \leq \frac{\rho\mu(\sigma_O^2 + \sigma_U^2)}{2(1 - \rho)} + \frac{1}{\mu}$$

$$\begin{aligned} \Leftrightarrow \frac{\rho(1+C_U^2)-2+2(1-\rho)}{2\mu(1-\rho)} &\leq W_s \leq \frac{\rho\mu^2(\sigma_O^2+\sigma_U^2)+2(1-\rho)}{2\mu(1-\rho)} \\ \Leftrightarrow \frac{\rho(C_U^2-1)}{2\mu(1-\rho)} &\leq W_s \leq \frac{\rho\mu^2(\sigma_O^2+\sigma_U^2)+2(1-\rho)}{2\mu(1-\rho)}. \end{aligned}$$

Finalmente, o número médio de clientes no sistema é

$$L_s = \lambda W_s,$$

logo

$$\frac{\rho^2(C_U^2-1)}{2(1-\rho)} \leq L_s \leq \frac{(\rho\mu)^2(\sigma_O^2+\sigma_U^2)+2(1-\rho)}{2(1-\rho)}.$$

Marshall [30] determinou um limite mais apertado para uma subclasse deste modelo, uma subclasse que engloba a maioria dos casos que se pode encontrar na prática. Esta subclasse engloba os sistemas que satisfazem a seguinte relação, que para todos os valores constantes $t_0 \geq 0$,

$$E[X - t_0 | X > t_0] \leq \frac{1}{\lambda} \quad (53)$$

ou seja, se soubermos que o tempo entre as chegadas demorou mais do que t_0 , então a duração prevista do tempo restante $X - t_0$ é menor ou igual a $\frac{1}{\lambda}$.

Quando a condição (53) é satisfeita, temos:

$$B - \frac{1+\rho}{2\lambda} \leq W_q \leq B,$$

onde B é o limite superior de (52), ou seja,

$$B = \frac{\lambda(\sigma_O^2 + \sigma_U^2)}{2(1-\rho)}. \quad (54)$$

Usamos agora as relações de Little, para o número médio de clientes no sistema, L_s , na fila, L_q , e o tempo médio esperado pelo cliente no sistema, W_s . Como

$$L_q = \lambda W_q$$

vem

$$\lambda B - \frac{1+\rho}{2} \leq L_q \leq \lambda B. \quad (55)$$

De

$$W_s = W_q + \frac{1}{\mu}$$

resulta

$$B + \frac{1}{\mu} - \frac{1+\rho}{2\lambda} \leq W_s \leq B + \frac{1}{\mu}.$$

E, finalmente

$$L_s = \lambda W_s$$

logo

$$\lambda B + \rho - \frac{1+\rho}{2} \leq L_s \leq \lambda B + \rho.$$

Estes limites são muito bons. De facto utilizando (55) e tendo em conta que $0 < \rho < 1$ obtemos que

$$1 < 1 + \rho < 2 \Rightarrow \frac{1}{2} < \frac{1 + \rho}{2} < 1,$$

ou seja, a diferença entre o limite inferior ao limite superior está sempre entre $\frac{1}{2}$ e 1. Isto implica que é possível determinar no número médio de clientes na fila com uma diferença de, no máximo, 1 cliente, dependendo de ρ .

Quando consideramos a v.a. exponencial na expressão (53) temos uma igualdade. De facto, se $X \sim \text{Exponencial}(\lambda)$ (como nos modelos $M/M/1$ e $M/G/1$) sabemos de (2) que

$$X - t_0 | X > t_0 \stackrel{d}{=} X$$

e, consequentemente,

$$E[X - t_0 | X > t_0] = E[X] = \frac{1}{\lambda}.$$

Existe um outro resultado, habitualmente designado por “aproximação de tráfego pesado” (cf. [23]) que nos dá uma aproximação dos valores anteriores quando os valores de ρ estão muito próximos de 1.

Este resultado indica que quando $\rho \approx 1$, a distribuição no estado de equilíbrio do tempo médio que um cliente espera na fila é aproximado a uma distribuição exponencial com valor médio

$$W_q \approx \frac{\lambda(\sigma_O^2 + \sigma_U^2)}{2(1 - \rho)}.$$

Esta expressão é idêntica ao limite superior da expressão (52), mostrando assim que o limite superior melhora quando $\rho \approx 1$.

Neste caso o número médio de clientes na fila é, aproximadamente,

$$\begin{aligned} L_q &= \lambda W_q \\ &\approx \frac{\lambda^2(\sigma_O^2 + \sigma_U^2)}{2(1 - \rho)}, \end{aligned}$$

o tempo médio de espera de um cliente no sistema é, aproximadamente,

$$\begin{aligned} W_s &= W_q + \frac{1}{\mu} \\ &\approx \frac{\lambda(\sigma_O^2 + \sigma_U^2)}{2(1 - \rho)} + \frac{1}{\mu} = \frac{\lambda(\sigma_O^2 + \sigma_U^2)}{2(1 - \rho)} + \frac{\rho}{\lambda} \\ &= \frac{\lambda^2(\sigma_O^2 + \sigma_U^2) + 2\rho(1 - \rho)}{2\lambda(1 - \rho)}, \end{aligned}$$

e o número médio de clientes no sistema é, aproximadamente,

$$\begin{aligned} L_s &= \lambda W_s \\ &\approx \frac{\lambda^2(\sigma_O^2 + \sigma_U^2) + 2\rho(1 - \rho)}{2(1 - \rho)}. \end{aligned}$$

O seguinte quadro reúne as principais características do modelo $G/G/1$,

Número médio de clientes no sistema	$\lambda B + \rho - \frac{1+\rho}{2} \leq L_s \leq \lambda B + \rho, B \text{ dado por (54)}$
Comprimento médio da fila	$\lambda B - \frac{1+\rho}{2} \leq L_q \leq \lambda B, B \text{ dado por (54)}$
Tempo médio de espera no sistema	$B + \frac{1}{\mu} - \frac{1+\rho}{2\lambda} \leq W_s \leq B + \frac{1}{\mu}, B \text{ dado por (54)}$
Tempo médio de espera na fila	$B - \frac{1+\rho}{2\lambda} \leq W_q \leq B, B \text{ dado por (54)}$
Número médio de clientes no sistema com $\rho \approx 1$	$L_s \approx \frac{\lambda^2(\sigma_O^2 + \sigma_U^2) + 2\rho(1-\rho)}{2(1-\rho)}$
Comprimento médio da fila com $\rho \approx 1$	$L_q \approx \frac{\lambda^2(\sigma_O^2 + \sigma_U^2)}{2(1-\rho)}$
Tempo médio de espera no sistema com $\rho \approx 1$	$W_s \approx \frac{\lambda^2(\sigma_O^2 + \sigma_U^2) + 2\rho(1-\rho)}{2\lambda(1-\rho)}$
Tempo médio de espera na fila com $\rho \approx 1$	$W_q \approx \frac{\lambda(\sigma_O^2 + \sigma_U^2)}{2(1-\rho)}$
Taxa de ocupação	$\rho = \frac{\lambda}{\mu}$

Quadro 14: Sumário das características do modelo $G/G/1$

5.7 Modelo $G/G/S$

Por último, vamos analisar o modelo $G/G/S$ que é idêntico ao modelo anterior, $G/G/1$, com a única diferença que contém S servidores em vez de um.

Brumelle [10] mostrou que

$$W_{q \text{ } G/G/1} - \frac{(S-1)\mu E[U^2]}{2S} \leq W_q \leq \frac{\left[\sigma_O^2 + \frac{1}{S}\sigma_U^2 + \frac{S-1}{(S\mu)^2}\right]\lambda}{2(1-\rho)}$$

isto é,

$$W_{q \text{ } G/G/1} - \frac{(S-1)\mu E[U^2]}{2S} \leq W_q \leq \frac{\lambda\sigma_O^2 + \mu\rho\sigma_U^2 + \rho\frac{S-1}{S\mu}}{2(1-\rho)}$$

onde $\rho = \frac{\lambda}{S\mu}$, σ_U^2 é a variância da taxa de serviço, σ_O^2 é a variância da taxa de chegada, $E[U^2]$ é o segundo momento do tempo de serviço para cada um dos S servidores e $W_{q \text{ } G/G/1}$ é o tempo médio que um cliente espera na fila no modelo $G/G/1$.

Usando as fórmulas de Little, obtemos o número médio de clientes na fila. De

$$L_q = \lambda W_q$$

vem

$$\begin{aligned} \lambda W_{q \text{ } G/G/1} - \lambda \frac{(S-1)\mu E[U^2]}{2S} &\leq L_q \leq \lambda \frac{\lambda\sigma_O^2 + \mu\rho\sigma_U^2 + \rho\frac{S-1}{S\mu}}{2(1-\rho)} \\ \Leftrightarrow L_{q \text{ } G/G/1} - \lambda \frac{(S-1)\mu E[U^2]}{2S} &\leq L_q \leq \frac{\lambda^2\sigma_O^2 + \mu\lambda\rho\sigma_U^2 + \rho^2(S-1)}{2(1-\rho)}. \end{aligned}$$

O tempo médio de espera de um cliente no sistema é

$$W_s = W_q + \frac{1}{\mu}$$

logo

$$\begin{aligned} W_{q \text{ } G/G/1} - \frac{(S-1)\mu E[U^2]}{2S} + \frac{1}{\mu} &\leq W_s \leq \frac{\lambda\sigma_O^2 + \mu\rho\sigma_U^2 + \rho\frac{S-1}{S\mu}}{2(1-\rho)} + \frac{1}{\mu} \\ \Leftrightarrow W_{s \text{ } G/G/1} - \frac{(S-1)\mu E[U^2]}{2S} &\leq W_s \leq \frac{\lambda\sigma_O^2 + \mu\rho\sigma_U^2 + \rho\frac{S-1}{S\mu}}{2(1-\rho)} + \frac{1}{\mu}. \end{aligned}$$

O número médio de clientes na fila é

$$L_s = \lambda W_s$$

logo

$$\begin{aligned} \lambda W_{s \text{ } G/G/1} - \lambda \frac{(S-1)\mu E[U^2]}{2S} &\leq L_s \leq \lambda \frac{\lambda\sigma_O^2 + \mu\rho\sigma_U^2 + \rho\frac{S-1}{S\mu}}{2(1-\rho)} + \lambda \frac{1}{\mu} \\ \Leftrightarrow L_{s \text{ } G/G/1} - \lambda \frac{(S-1)\mu E[U^2]}{2S} &\leq L_s \leq \frac{\lambda^2\sigma_O^2 + \mu\lambda\rho\sigma_U^2 + \rho^2(S-1)}{2(1-\rho)} + \rho S. \end{aligned}$$

Tal como foi feito no modelo anterior, podemos usar o resultado das “aproximações de tráfego pesado”, segundo [24], quando $\rho \approx 1$ temos uma aproximação à distribuição exponencial com o tempo médio de espera de um cliente na fila é, aproximadamente,

$$W_q \approx \frac{\lambda \left(\sigma_O^2 + \frac{\sigma_U^2}{S} \right)}{2(1-\rho)}.$$

ou, equivalentemente,

$$W_q \approx \frac{\rho S \mu \sigma_O^2 + \rho \mu \sigma_U^2}{2(1-\rho)} = \rho \mu \frac{S \sigma_O^2 + \sigma_U^2}{2(1-\rho)}$$

Vamos utilizar como anteriormente as fórmulas de Little. O número médio de clientes na fila é, aproximadamente,

$$\begin{aligned} L_q &= \lambda W_q \\ &\approx \lambda \rho \mu \frac{S \sigma_O^2 + \sigma_U^2}{2(1-\rho)}, \end{aligned}$$

o tempo médio que um cliente tem de esperar no sistema é, aproximadamente,

$$\begin{aligned} W_s &= W_q + \frac{1}{\mu} \\ &\approx \rho \mu \frac{S \sigma_O^2 + \sigma_U^2}{2(1-\rho)} + \frac{1}{\mu}, \end{aligned}$$

e o número médio de clientes no sistema é, aproximadamente,

$$\begin{aligned} L_s &= \lambda W_s \approx \lambda \rho \mu \frac{S \sigma_O^2 + \sigma_U^2}{2(1-\rho)} + \frac{\lambda}{\mu} \\ &= \lambda \rho \mu \frac{S \sigma_O^2 + \sigma_U^2}{2(1-\rho)} + S \rho. \end{aligned}$$

O seguinte quadro reúne as principais características deste modelo,

Número médio de clientes no sistema	$L_s \text{ } G/G/1 - \lambda \frac{(S-1)\mu E[U^2]}{2S} \leq L_s \leq \frac{\lambda^2 \sigma_O^2 + \mu \lambda \rho \sigma_U^2 + \rho^2 (S-1)}{2(1-\rho)} + \rho S$
Comprimento médio da fila	$L_q \text{ } G/G/1 - \lambda \frac{(S-1)\mu E[U^2]}{2S} \leq L_q \leq \frac{\lambda^2 \sigma_O^2 + \mu \lambda \rho \sigma_U^2 + \rho^2 (S-1)}{2(1-\rho)}$
Tempo médio de espera no sistema	$W_s \text{ } G/G/1 - \frac{(S-1)\mu E[U^2]}{2S} \leq W_s \leq \frac{\lambda \sigma_O^2 + \mu \rho \sigma_U^2 + \rho \frac{S-1}{S\mu}}{2(1-\rho)} + \frac{1}{\mu}$
Tempo médio de espera na fila	$W_q \text{ } G/G/1 - \frac{(S-1)\mu E[U^2]}{2S} \leq W_q \leq \frac{\lambda \sigma_O^2 + \mu \rho \sigma_U^2 + \rho \frac{S-1}{S\mu}}{2(1-\rho)}$
Número médio de clientes no sistema com $\rho \approx 1$	$L_s \approx \lambda \rho \mu \frac{S \sigma_O^2 + \sigma_U^2}{2(1-\rho)} + S \rho$
Comprimento médio da fila com $\rho \approx 1$	$L_q \approx \lambda \rho \mu \frac{S \sigma_O^2 + \sigma_U^2}{2(1-\rho)}$
Tempo médio de espera no sistema com $\rho \approx 1$	$W_s \approx \rho \mu \frac{S \sigma_O^2 + \sigma_U^2}{2(1-\rho)} + \frac{1}{\mu}$
Tempo de espera na fila com $\rho \approx 1$	$W_q \approx \rho \mu \frac{S \sigma_O^2 + \sigma_U^2}{2(1-\rho)}$
Taxa de ocupação	$\rho = \frac{\lambda}{S\mu}$

Quadro 15: Sumário das características do modelo $G/G/S$

6 Conclusão

Ao longo desta tese foram estudados alguns sistemas de filas de espera. Muito mais se poderia dizer, pois este tema é tão vasto e este trabalho é uma breve introdução com alguns resultados. Por exemplo, existem estudos feitos para o caso em que os clientes chegam em grupo (o que acontece por exemplo na entrada de um parque temático), para o caso em que os clientes são atendidos em grupo (o que acontece por exemplo quando pensamos no serviço prestado por uma carreira de autocarros), para o caso em que os clientes são impacientes e desistem enquanto esperam na fila, para o caso em que há clientes prioritários (num hospital, por exemplo), entre outros. Esperamos, no entanto, ter conseguido dar a conhecer o essencial desta teoria e ter despertado o interesse para a exploração de modelos mais complexos.

Referências

- [1] Abad, R. (2002), *Introducción a la simulación y a la teoría de colas*, Netbiblo, SL.
- [2] Adan, I. e Resing, J. (2001), *Queueing Theory*, J. C. Baltzer AG, Science Publishers, Vol. 37, No. 1/3, pp. 65-98.
- [3] Adan, I., Waarsenburg, W. e Wessels, J. (1996), Analyzing $E_k/E_r/c$ queues, *European Journal of Operational Research*, Vol. 92, No. 1, pp. 112-124.
- [4] Adan, I. e Zhao, Y. (1996), Analyzing $GI/E_r/1$ queues, *Operations Research Letters*, Vol. 19, No. 4, pp. 183-190.
- [5] Allen, A. (1978), *Probability, Statistics, and Queueing Theory with Computer Science Applications*, Academic Press.
- [6] Beichelt, F. (2006), *Stochastic Processes in Science, Engineering and Finance*, Chapman & Hall/CRC.
- [7] Benes, V. (1963), *General Stochastic Processes in the Theory of Queues*, Addison-Wesley.
- [8] Bose, S. (2002), *An introduction to queueing systems*, Kluwer Academic/ Plenum Publishers.
- [9] Brockmeyer, E., Halstrøm e Jensen, A. (1948), The life and works of A. K. Erlang, *Transactions of the Danish Academy of Technical Sciences*, No. 2, pp. 1-288.
- [10] Brumelle, S. (1971), Some Inequalities for Parallel-Server Queues, *Operations Research*, Vol. 19, No. 2, pp. 402-413.
- [11] Cohen, J. (1969), *The Single Server Queue*, North Holland.
- [12] Cox, D. (1955), The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables, *Proc. Camb. Phil. Soc.*, Vol. 50, pp. 433-441.
- [13] Cox, D. e Smith, W. (1961), *Queues*, Methuen.
- [14] Çinlar, E. (1975), *Introduction to stochastic processes*, Prentice-Hall, Inc.
- [15] Disney, R. (1986), The making of a queueing theorist, *The Craft of Probability Modelling*, Springer-Verlag, pp. 196-212.
- [16] Feller, W. (1957), *An Introduction to Probability Theory and Its Applications*, Vol. 1, Wiley.
- [17] Feller, W. (1957), *An Introduction to Probability Theory and Its Applications*, Vol. 2, Wiley.
- [18] Gnedenko, B. e Kovalenko, I. (1989), *Introduction to Queueing Theory*, Birkhäuser.
- [19] Hillier, F. e Lieberman, G. (1995), *Introduction to Operations Research*, Mc Graw-Hill.
- [20] Jaiswal, N. (1968), *Priority Queues*, Academic Press.
- [21] Kendall, D. (1951), Some problems in the theory of queues, *J.R. Statist. Soc.*, Vol. 13, pp. 151-185.

- [22] Kendall, D. (1953), Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain, *Ann. Math. Statist.*, Vol. 24, pp. 338-354.
- [23] Kingman, J. (1962), On Queues in Heavy Traffic, *Journal of the Royal Statistical Society, Series B*, Vol. 24, pp. 383-392.
- [24] Kollerstrom, L. (1974), Heavy Traffic Theory for Queues with Several Servers: I, *Journal of Applied Probability*, Vol. 11, pp. 544-552.
- [25] Larson, R. e Odoni, A. (1981), *Urban Operations Research*, Prentice-Hall (http://web.mit.edu/urban_or_book/www/book/index.html).
- [26] Lindley, D. (1952), The theory of queues with a single server, *Proc. Camb. Phil. Soc.*, Vol. 48, pp. 277-289.
- [27] Loynes, R. (1962), The stability of a queue with non-independent inter-arrival and service times, *Proc. Camb. Phil. Soc.*, Vol. 58, pp. 497-520.
- [28] Loynes, R. (1962), Stationary waiting-time distributions for single-server queues, *Ann. Math. Statist.*, Vol. 33, No. 4, pp. 1321-1339.
- [29] Marchal, G. (1978), Some Simpler Bounds on the Mean Queueing Time, *Operations Research*, Vol. 26, pp. 1083-1088.
- [30] Marshall, T. (1968), Some Inequalities in Queueing, *Operations Research*, Vol. 16, pp. 651-665.
- [31] Morse, P. (1958), *Queues, Inventories, and Maintenance*, Wiley.
- [32] Müller, D. (2007), *Processos Estocásticos e Aplicações*, Almedina.
- [33] Pestana, D. e Velosa, S. (2002), *Introdução à Probabilidade e à Estatística*, Calouste Gulbenkian.
- [34] Prabhu, N. (1965), *Queues and Inventories: A Study of Their Basic Stochastic Processes*, Wiley.
- [35] Ravidran, A. (2007), *Operations Research and Management Science Handbook*, CRC Press, Taylor & Francis Group.
- [36] Resnick, S. (1992), *Adventures in Stochastic Processes*, Birkhäuser.
- [37] Saaty, T. (1961), *Elements of Queueing Theory*, Mc Graw-Hill.
- [38] Smith, W. (1953), Regenerative stochastic processes, *Proc. Roy. Soc.*, Vol. 232, pp. 6-31.
- [39] Stidham, S. Jr. (2002), *Applied Probability in Operations Research: A Retrospective*, Chapel Hill.
- [40] Syski, R. (1960), *Introduction to Congestion Theory in Telephone Systems*, Oliver and Boyd.
- [41] Takagi, H. (1991), *Queueing Analysis: A Foundation of Performance Evaluation, Volume I — Vacation and Priority Systems*, North - Holland, The Netherlands.
- [42] Taylor, H. e Karlin, S. (1998), *An Introduction To Stochastic Modeling*, Academic Press.
- [43] Whittle, P. (2001), Applied probability in Great Britain, *Operations Research*, Vol. 50, No. 1, pp. 227-239.

Índice

- acontecimento, 4
 - com independência, 5
 - complementar, 4
- capacidade do sistema, 18
- características das filas de espera, 20
- chegadas, 16
 - atitude dos clientes, 17
 - controle das, 16
 - dimensão da, 16
 - dimensão da população, 16
 - distribuição das, 17
 - taxa de, 17
- covariância, 9
- disciplina de atendimento, 18
 - FCFS ou FIFO, 18
 - GD, 19
 - LCFS ou LIFO, 18
 - PRI, 18
 - RR, 19
 - SIRO, 18
- espaço amostral, 4
- estado de equilíbrio, 21
- experiência aleatória, 4
- fórmulas de Little, 21
- fila de espera, 17
 - comprimento da, 17
 - disciplina da, 17
 - número de, 17
- fórmula de perda de Erlang, 63, 94
- função de distribuição, 6
 - de Erlang, 10
 - de Poisson, 9
 - exponencial, 9
 - gama, 10
- função densidade de probabilidade, 7
- função massa de probabilidade, 6
- função mensurável, 4
- função geradora de probabilidades, 8
- $G/G/1$, 99
- $G/G/S$, 103
- $G/M/1$, 95
- $M/E_r/1$, 80
- $M/G/1$, 85
- $M/G/\infty$, 93
- $M/G/S/S$, 94
- $M/M/1$, 23
- $M/M/1/\infty/H$, 66
- $M/M/1/K$, 44
- $M/M/\infty$, 41
- $M/M/S$, 32
- $M/M/S/\infty/H$, 72
- $M/M/S/K$, 53
- $M/M/S/S$, 62
- medida, 5
 - de probabilidade, 5
- modelo markoviano, 23
- modelo não markoviano, 80
- momento de ordem n , 7
- probabilidade condicionada, 5
- probabilidade do estado de transição, 14
- probabilidade total, 5
- processo de morte, 13
- processo de nascimento, 12
- processo de nascimento e morte, 14
 - distribuição estacionária do , 14
- processo estocástico, 11
 - de Markov, 11
 - de Poisson, 11
- serviço, 17
 - configuração de, 17
 - dimensão de, 18
 - distribuição do tempo de, 18
 - taxa de, 18
- sigma-álgebra, 4
- transformada de Laplace, 8
- valor médio, 7
- variável aleatória, 6
 - absolutamente contínua, 6
 - com independência, 7
 - contínua, 6
 - discreta, 6
- variância, 8
- variantes das filas de espera, 19

